**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, 1/70 — A1

(11) International Publication Number: WO 95/01453

(43) International Publication Date: 12 January 1995 (12.01.95)

(21) International Application Number: PCT/US94/07416

(22) International Filing Date: 30 June 1994 (30.06.94)

(30) Priority Data:
08/087,010 — 1 July 1993 (01.07.93) — US
08/241,373 — 11 May 1994 (11.05.94) — US

(71) Applicants: THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY [US/US]; Stanford, CA 94305 (US). THE GOVERNMENT OF THE UNITED STATES OF AMERICA, represented by THE SECRETARY OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES AND HIS SUCCESSORS [US/US]; Washington, DC 20231 (US).

(72) Inventors: MULLINS, James, Ivan; 262 Hillsdale Way, Redwood City, CA 94062-3928 (US). DELWART, Eric, Lawrence; 2158 Williams, Palo Alto, CA 94301 (US).

(74) Agent: FABIAN, Gary, R.; Dehlinger & Associates, P.O. Box 60850, Palo Alto, CA 94306-0850 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

**Published**

*With international search report.*

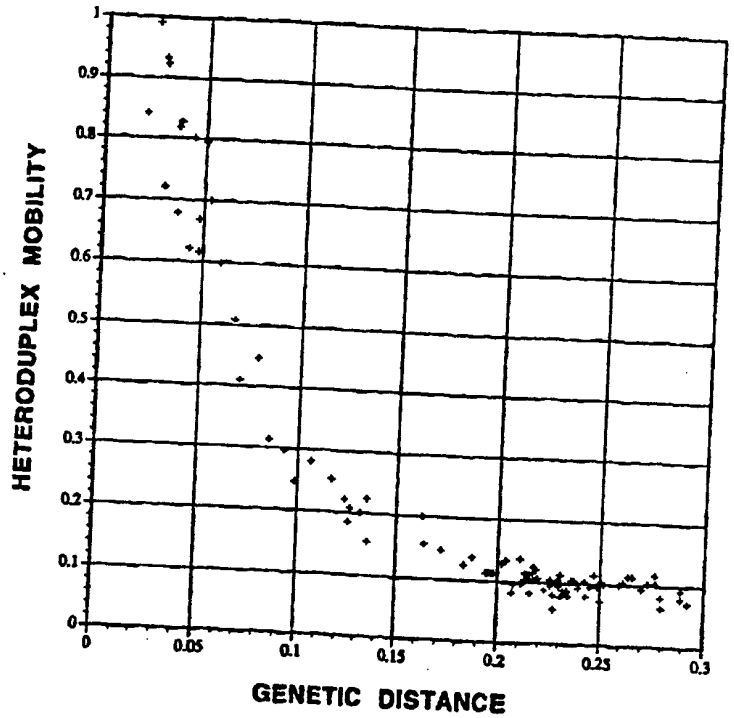
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

Example A provides a method of assaying a heteroduplex mobility assay

(54) Title: A HETERODUPLEX MOBILITY ASSAY FOR THE ANALYSIS OF NUCLEIC ACID SEQUENCE DIVERSITY

(57) Abstract

Heteroduplexes formed between different members of a gene family migrate more slowly in non-denaturing polyacrylamide gels (PAGE) than homoduplexes. Heteroduplexes with deletions/insertions as small as a single nucleotide are identified by a mobility retardation in PAGE. Heteroduplexes containing only mismatches also display a mobility shift whose magnitude depends on the % sequence divergence. Mobility shifts of deletion/insertion heteroduplexes are affected by the size, location and number of deletions/insertions and by the sequence of the unpaired and neighboring mismatched nucleotides. The heteroduplex mobility assay (HMA) and heteroduplex tracking assay (HTA) of the present invention can be used for tracking viral quasispecies evolution within and between individuals, for tracking species of infectious agents, and for monitoring DNA sequence changes including identification and tracking of cellular DNA polymorphisms.



*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| **AT** | Austria | **GB** | United Kingdom | **MR** | Mauritania |
| **AU** | Australia | **GE** | Georgia | **MW** | Malawi |
| **BB** | Barbados | **GN** | Guinea | **NE** | Niger |
| **BE** | Belgium | **GR** | Greece | **NL** | Netherlands |
| **BF** | Burkina Faso | **HU** | Hungary | **NO** | Norway |
| **BG** | Bulgaria | **IE** | Ireland | **NZ** | New Zealand |
| **BJ** | Benin | **IT** | Italy | **PL** | Poland |
| **BR** | Brazil | **JP** | Japan | **PT** | Portugal |
| **BY** | Belarus | **KE** | Kenya | **RO** | Romania |
| **CA** | Canada | **KG** | Kyrgystan | **RU** | Russian Federation |
| **CF** | Central African Republic | **KP** | Democratic People's Republic of Korea | **SD** | Sudan |
| **CG** | Congo | | | **SE** | Sweden |
| **CH** | Switzerland | **KR** | Republic of Korea | **SI** | Slovenia |
| **CI** | Côte d'Ivoire | **KZ** | Kazakhstan | **SK** | Slovakia |
| **CM** | Cameroon | **LI** | Liechtenstein | **SN** | Senegal |
| **CN** | China | **LK** | Sri Lanka | **TD** | Chad |
| **CS** | Czechoslovakia | **LU** | Luxembourg | **TG** | Togo |
| **CZ** | Czech Republic | **LV** | Latvia | **TJ** | Tajikistan |
| **DE** | Germany | **MC** | Monaco | **TT** | Trinidad and Tobago |
| **DK** | Denmark | **MD** | Republic of Moldova | **UA** | Ukraine |
| **ES** | Spain | **MG** | Madagascar | **US** | United States of America |
| **FI** | Finland | **ML** | Mali | **UZ** | Uzbekistan |
| **FR** | France | **MN** | Mongolia | **VN** | Viet Nam |
| **GA** | Gabon | | | | |

# A HETERODUPLEX MOBILITY ASSAY FOR THE ANALYSIS OF NUCLEIC ACID SEQUENCE DIVERSITY

## Field of the Invention

The present invention describes a method for genetic analysis of gene families: the Heteroduplex Mobility Assay (HMA). The method has utility in the analysis of gene pools, in particular, gene pools of disease causing microorganisms. Further, the present invention describes a method for evaluating the effects of a disease treatment protocol on DNA sequence variation of a nucleic acid target sequence associated with the disease, using the Heteroduplex Mobility Assay (HMA) or Heteroduplex Tracking Assay (HTA). The method has utility in evaluating sequence variation among gene pools of disease causing microorganisms and variant disease-related genes, such as oncogenes.

## References

Adachi, A., *et al.*, *J. Virol.* 59:284-291 (1986).

Ausubel, F. M., *et al.*, Current Protocols in Molecular Biology, John Wiley and Sons, Inc., Media PA.

Bayever, E., *et al.*, *Antisense Research and Development* 3:383-390 (1993).

Borresen, *et al.*, *PNAS* 88(19):8405-9 (1991).

Bowman, *et al.*, *Mol. Biol. and Evolution* 9:893-904 (1992).

Breniere, *et al.*, *Am. J. Trop. Med. and Hygiene* 46:335-41 (1992).

Briones, *et al.*, *Mol. and Bio. Parasitology* 53:121-7 (1991).

Calabretta, B., *et al.*, *Seminars in Cancer Biol.* 3(6):391-398 (1992).

Calabretta, B., *et al.*, *Cancer Treatment Rev.* 19(2):169-179 (1993).

Clark, *et al.*, *N. Engl. J. Med.* 324:954-60 (1991).

Cotton, *et al.*, *PNAS* 85:4397-4401 (1988).

Daar, E.S., *et al.*, *Proc. Natl. Acad. Sci. USA* 87:6574-6578 (1990).

Daar, *et al.*, *N. Eng. J. Med.* 32:961-964 (1991).

Ellis, *et al.*, *Mol. and Bio. Parasitology* 54:87-95 (1992).

Felsenstein, J., *Ann. Rev. Genet.* 22:521-65 (1988).

Felsenstein, J., *Cladistics* 5:164-166 (1989).

Folks, T.M., *et al.*, *Proc. Natl. Acad. Sci. USA* 86:2365-2368 (1989).

Fujita, K., *et al.*, *J. Virol.* 66:4445-4451 (1992).

Gazdar, A. F., *et al.*, U.S. Patent No. 4,892,829, issued Jan. 09, 1990.

Ghossein, R.A., *et al.*, Diagnostic Mol. Pathol. 1(3):185-191 (1992).

Kuiken, C.L., *et al.*, *J. Virol.* 66:4622-4627 (1992).

Kleemola, *et al.*, *Pediatric Infect. Dis. J.* 12:344-5 (1993).

Kusumi, K., *et al.*, *J. Virol.* 66:875-885 (1992).

Larder, B.A., *et al.*, *Science* 246:1155-1158 (1989).

Lopez-Galindez, C., *et al.*, *Proc. Natl. Acad. Sci. USA* 88:4280-4284 (1991).

Maniatis, T., *et al.* Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory (1982).

McCutchan, F.E., *et al.*, *J. Acquir. Immune Defic. Syndr.* 4:1241-1250 (1991).

McCutchan, F.E., *et al.*, *J. Acquir. Immune Defic. Syndr.* 5:441-449 (1992).

Meyerhans, A., *et al.*, *Cell* 58:901-910 (1989).

Mullis, K., U.S. Patent No. 4,683,202, issued July 28, 1987.

Myers, G., *et al.*, *Nature* 313:495 (1985a).

Myers, G., *et al.*, *Science* 230:1242 (1985b).

Myers, G., et al, Evolutionary Potential of Complex Retroviruses, (Wagner, *et al.*, Eds.), Plenum Press, New York (1991).

Myers, G., *et al.*, *AIDS Res. & Human Retroviruses* 8:373-386 (1992a).

Myers, G., *et al.*, Human Retroviruses and AIDS, Los Alamos National Laboratory (1992b).

Myers, R.M., *et al.*, *Nuc. Acids Res.* 13:3111-3128 (1985a).

Myers, R.M., et al., *Nuc. Acids Res.* 13:3131-3145 (1985b).

Myers, *et al.*, Methods in Enzymology Vol. 155, pp 501-526, Academic Press Inc. (1987).

Orita, *et al.*, *PNAS* 86:2766-2770 (1989).

Ou, C.-Y., *et al.*, *Science* 256:1165-1171 (1992).

Palca, *et al.*, *Science* 256:1387-1388 (1992).

Ratner, L., *et al.*, *Nature* 313:277-284 (1985).

Sambrook, J., et al., In Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Vol. 2 (1989).

Sheffield, *et al.*, *PNAS* 86:232-236 (1989).

Simmonds, P.P., *et al.*, *J. Virol.* 65:6266-6276 (1991).

Simmonds, P., *et al.*, *J. Virol.* 64:864-872 (1990a).

Simmonds, P., *et al.*, *J. Virol.* 64:5840-5850 (1990b).

St. Clair, M.H., *et al.*, *Science* 253:1557-1559 (1991).

Tersmette, M., *et al., J. Virol.* 63:2118-2125 (1989).

Vohra, *et al., J. Mol. Eval.* 34:303-395 (1992).

Weisburg, W.G., *et al., J. Bacteriology* 171(12):6455-6467 (1989).

Weiss, *et al., Mol. & Bioch. Parasitology* 54:72-86 (1992).

Wickstrom, E., Editor, Prospects for Antisense Nucleic Acid Therapy of Cancer and AIDS, Wiley-Liss, New York, NY (1991).

Wolfe, K.H., *et al.*, Plant Molec. Biol. 18(6):1037-1048 (1992).

Zalewski, A., *et al., Circulation Res.* 88:1190-1195 (1993).

Zimmerman, *et al., Mol. & Bioch. Parasitology* 58:259-267 (1993).

**Background of the Invention**

The ability to detect small changes in DNA sequences, such as, base substitutions, deletions, and insertions, has become important in a number of applications including diagnosis of human genetic diseases. Several methods for the detection of small sequence changes have been proposed.

One method involves the use of restriction fragment polymorphism (RFLP) analysis. Restriction fragment polymorphisms arise when a base change results in the loss or acquisition of a restriction endonuclease cleavage site in a defined DNA sample, for example, a selected region of a viral genome. However, since the chances are low that a particular base change will eliminate or create a restriction site, RFLP analysis is labor intensive and tends not to provide a high

level of sensitivity since most changes in a DNA sequence go undetected.

Two methods for the detection and localization of base changes have been described by Myers, *et al.* (1987). In both methods, a radioactively labeled single-strand wild-type sequence probe is annealed to a test sample of cloned or genomic DNA. If the sample DNA contains an base change relative to the wild-type sequence, then a mismatch is formed at that site. In the first method, a DNA or RNA probe is used and mismatched duplexes are separated from perfectly paired duplexes by denaturing gradient gel electrophoresis (Myers, *et al.*, 1985a). In the second method, the probe is a single-strand RNA molecule having a defined sequence. In this method ribonuclease is used to cleave any mismatched duplexes. Any cleaved products are then identified by polyacrylamide gel electrophoresis and autoradiography (Myers, *et al.*, 1985b).

Polymerase chain reaction has provided the means to obtain large scale amplification of target DNA (Mullis; Mullis, *et al.*) from, for example, defined genomic regions of DNA. A number of approaches have been utilized, for the detection of DNA mutations, in combination with polymerase chain reaction amplified products. These approaches include denaturing polyacrylamide gel electrophoresis (see above and Sheffield, *et al.*, 1989), single-chain conformation polymorphism analysis (Orita, *et al.*), chemical cleavage of mismatches (Cotton, *et al.*). Constant denaturant gel electrophoresis has also been employed as a screening method for mutations in a defined DNA sequence (Borresen, *et al.*, 1991)

**Summary of the Invention**

The present invention provides a method of evaluating sequence diversity in a mixture of nucleic acids containing a target sequence. In the method of the present invention, amplification primers are selected which are complementary to nucleic acid sequences flanking the target region. The nucleic acids and the primers are combined under conditions that promote the hybridization of the primers to the nucleic acids, thus generating primer/nucleic acid complexes. These complexes are converted to double-strand fragments in the presence of a suitable polymerase and all four deoxyribonucleotides. The primer-containing fragments are amplified by repeated rounds of primer extension until a desired degree of amplification has been achieved. The resulting double-stranded amplified fragments are denatured and renatured to form a population of amplified fragment DNA duplexes. These duplexes are separated on polyacrylamide gels. The relative migration of the duplexes is analyzed to establish the relative degree of sequence relatedness in the population of amplified fragments.

In the method, the starting nucleic acids can be RNA molecules which are converted to DNA templates using reverse transcriptase.

Typically, the denaturing step of the method is thermal denaturing, and double-stranded fragments are generated using a thermostable DNA polymerase.

In one embodiment, the amplified duplexes are separated by polyacrylamide gel electrophoresis (PAGE), and analyzed by visualization of the amplification products with ethidium bromide staining or autoradiography.

In another embodiment, the amplified duplexes are separated by PAGE. The locations of the duplexes are then analyzed by transferring the nucleic acids from the gel to a support membrane, and hybridizing the nucleic acid transferred to the membrane with a labelled probe specific for the desired amplification products.

In the present invention, the primers may contain at least one detection moiety. The detection moiety can be, but is not limited to, one or more of the following: a radioactive moiety, biotin, digoxigenin, or a chemiluminescent moiety.

The method of the present invention can be applied to the analysis of nucleic acid from any infectious agent, including viruses, bacteria, mycoplasma, parasites, and fungi. One exemplary application of the present method is in evaluating intra- and inter-patient sequence diversity of Human Immunodeficiency Virus (HIV). Further, the present method can be applied to the analysis of sequence variation between defined genetic loci, such as, oncogenes, protooncogenes, and disease-associated loci (*e.g.*, Duchenne's muscular dystrophy). This method is also useful to evaluate sequence diversity over time in mixtures of nucleic acids containing a target sequence, where the mixtures are serially obtained from a single source, for example, the single source may be a patient infected with HIV-1.

The present invention also includes a method of evaluating sequence diversity between two different sample mixtures of nucleic acids, where the nucleic acids contain a target sequence. In this method, the two samples are treated as described above. After amplification, the two amplified samples are mixed, denatured, and

renatured. The resulting duplex molecules are separated on polyacrylamide gels. The relative migration of the duplexes is then analyzed to establish the relative degree of sequence relatedness among the amplified fragments of the population. This method can be applied to the analysis of samples from different geographic locations, different patients, or for the same patient with different samples collected over time.

In one embodiment of this method, the amplified fragments of one sample nucleic acid are labelled with a detection moiety and the labeled fragments are mixed with a molar excess of the amplified fragments of the other sample.

Yet another embodiment of the present invention is a method for detecting the presence of a selected nucleic acid target region in a nucleic acid sample. In this method a duplex DNA probe having two complementary strands is selected, where the duplex is homologous to the target region and each strand contains a detection moiety. Amplification primers complementary to nucleic acid sequences flanking the target region of the nucleic acid are also selected. The target region is amplified. The amplified products are denatured and mixed, in molar excess, with the duplex probe. The mixture is denatured, renatured, and analyzed as described above to establish the relative degree of sequence relatedness between the probe and sample target regions.

Typically, the mixing ratio for amplified fragments to probe is 100:1, amplified fragments to probe.

The method of the present invention has been described with reference to using amplification to

obtain samples suitable for heteroduplex mobility analysis. However, other methods of obtaining nucleic acid samples can be used as well, including, but not limited to, cloning sequences of interest or isolating particular restriction endonuclease digestion products from genomic DNA preparations.

Further, the present invention provides a method of evaluating, in mixtures of nucleic acids, the effect over time of a disease treatment, on DNA sequence variation of a nucleic acid target sequence associated with the disease. In the method of the present invention, amplification primers are chosen that are complementary to nucleic acid sequences flanking the target sequence. Mixtures of nucleic acids are serially obtained from a single source. For example, sera is collected from an HIV-infected individual before treatment and at selected time points over the course of treatment.

The nucleic acids from each sample and the primers are combined under conditions that promote the hybridization of the primers to the nucleic acids, thus generating primer/nucleic acid complexes. For each sample, these complexes are converted to double-strand fragments in the presence of a suitable polymerase and all four deoxyribonucleotides. The primer-containing fragments are amplified by repeated rounds of primer extension until a desired degree of amplification has been achieved.

The resulting double-stranded amplified fragments are denatured and renatured to form a population of amplified fragment DNA duplexes for each sample. These migration of the duplexes on the gel is analyzed for each sample to determine the relative migration of the duplexes and

establish the relative degree of sequence relatedness in the population of amplified fragments. The effect of the treatment is evaluated by comparing the relative degree of sequence relatedness of amplified fragments in each serial sample between the serial samples. Particularly, comparing to the pre-treatment, or zero time point.

In the method, the starting nucleic acids can be RNA molecules which are converted to DNA templates using reverse transcriptase.

Typically, the denaturing step of the method is thermal denaturing, and double-stranded fragments are generated using a thermostable DNA polymerase.

In one embodiment, the amplified duplexes are separated by polyacrylamide gel electrophoresis (PAGE), and analyzed by visualization of the amplification products with ethidium bromide staining or autoradiography.

In another embodiment, the amplified duplexes are separated by PAGE. The locations of the duplexes are then analyzed by transferring the nucleic acids from the gel to a support membrane, and hybridizing the nucleic acid transferred to the membrane with a labelled probe specific for the desired amplification products.

In the present invention, the primers may contain at least one detection moiety. The detection moiety can be, but is not limited to, one or more of the following: a radioactive moiety, biotin, digoxigenin, or a chemiluminescent moiety.

The method of the present invention can be applied to the analysis of nucleic acid from any microorganism/infectious agent, including viruses, bacteria, mycoplasma, mycobacteria, parasites, and

fungi. One exemplary application of the present method is in evaluating intra- patient sequence diversity of Human Immunodeficiency Virus (HIV).

Further, the present method can be applied to the analysis of sequence variation between defined genetic loci, such as, oncogenes, protooncogenes, and disease-associated loci (*e.g.*, Duchenne's muscular dystrophy).

Yet another embodiment of the present invention is a method for evaluating the effect of a disease treatment procedure on the presence of a selected nucleic acid target region in a nucleic acid sample. In this method a duplex DNA probe having two complementary strands is selected, where the duplex is homologous to the target region and each strand contains a detection moiety.

Serial samples are obtained from a single source, *e.g.*, a first sample before treatment and a second sample after treatment. Amplification primers complementary to nucleic acid sequences flanking the target region of the nucleic acid are also selected. The target region is amplified in each sample. The amplified products from each sample are denatured and mixed, in molar excess, with the duplex probe. Each mixture is denatured, renatured, and analyzed as described above to establish the relative degree of sequence relatedness between the probe and sample target regions.

Typically, the mixing ratio for amplified fragments to probe is 100:1, amplified fragments to probe.

The method of the present invention has been described with reference to using amplification to obtain samples suitable for heteroduplex mobility analysis. However, other methods of obtaining

nucleic acid samples can be used as well, including, but not limited to, cloning sequences of interest or isolating particular restriction endonuclease digestion products from genomic DNA preparations.

These and other objects and features of the invention will be more fully appreciated when the following detailed description of the invention is read in conjunction with the accompanying examples and drawings.

**Brief Description of the Figures**

Figure 1B shows the result of heteroduplex mobility (HIV-1 envelope gene DNA) analyzed on a 2.5% agarose gel. Figure 1A shows the same heteroduplex reactions separated on a 5% polyacrylamide gel. The heteroduplex samples were obtained from nested PCR amplifications performed using uncultured HIV-1 seropositive PBMC DNA.

Figures 2A to 2E show the results of heteroduplex analysis carried out with gap- and base-pair-mismatch-containing heteroduplex molecules. The heteroduplexes were resolved on agarose (2A) and polyacrylamide (2B-2E) gels.

Figure 3A shows a plot of the mobility shifts of heteroduplexes against the % divergence, based on nucleotide sequence analysis, between the two strands of the heteroduplexes. Figure 3B shows some of the heteroduplex mobility assay data used to generate the plot presented in Figure 3A.

Figure 4 shows the results of a heteroduplex mobility assay analysis of variant sequences from different HIV-1 quasispecies. In the figure, MA is HIV-positive asymptomatic, PE and BU were HIV-positive and symptomatic.

Figure 5 shows polyacrylamide heteroduplex mobility of inter-individual DNA heteroduplexes

using complex DNA mixtures (quasispecies) as probe and target. In the figure, "*" is unreannealed single stranded DNA. US quasispecies (1, 2, and 3) are from epidemiologically unrelated Americans and quasispecies 4 is from a patient from Zimbabwe.

Figure 6 shows polyacrylamide heteroduplex mobility of intra-individual DNA heteroduplexes using complex DNA mixtures (quasispecies) as probe and target. M1 and M27 represent probes derived from the month 1 and 27 quasispecies.

Figures 7A, 7B and 7C show polyacrylamide heteroduplex mobility of intra-individual DNA heteroduplexes using subclones derived from uncultured PBMC as probes and quasispecies as target.

Figure 8 shows a plot of the mobility shifts of heteroduplexes against the relative genetic distance between the two strands of the heteroduplexes.

Figures 9A and 9B show the results of HMA-based and DNA sequence-based estimates of genetic relatedness for variants of the HIV-*env* gene.

**Detailed Description of the Invention**

Definitions:

"Homoduplex molecules" are typically composed of two complementary DNA strands, where the strands have at least about 98% sequence homology.

"Heteroduplex molecules" are typically composed of two complementary DNA strands, where the strands have less than at least about 98% sequence homology. The functional definition of homoduplex and heteroduplex molecules, in the context of the present invention, is apparent from the results presented below. Typically, in a mixed population of homoduplex and heteroduplex

molecules, homoduplex molecules form a single band at a defined position in a polyacrylamide gel and heteroduplex molecules appear as slower migrating bands relative to the homoduplex.

"Gaps" occur when in duplexes consisting of two complementary DNA strands, where the first strand of the DNA contains more nucleotides at an internal site than the second strand DNA molecule, and where these extra nucleotides are flanked by paired-complementary sequences. Gaps can occur in heteroduplexes.

"Base-pair mismatches" typically refers to a single base-pair mismatch flanked by matched base-pairs. Base-pair mismatches also include a series of mismatched base-pairs flanked by matched base-pairs. Base-pair mismatches can occur in heteroduplexes.

"Quasispecies" refers to the totality of members of a species whose genomes have sequence divergence, but where the members still fall within the same species: for example, outgrowth of "a number of sequence variants" (the quasispecies) of HIV-1 from one major infecting variant.

"Microorganism" in the context of the present invention includes, but is not limited to, the following groups: bacteria, viruses, fungi, parasites and mycoplasma.

"Oncogene" includes genes that induce cancer or other uncontrolled proliferations of cells. Oncogenes can be mutated or activated (i) proto-oncogenes, or (ii) tumor suppressor genes (both of cellular origin), associated with the development and/or proliferation of tumor cells. Oncogenes, or portions thereof, may also be of viral origin.

I. Observations on Heteroduplex Mobility in Non-Denaturing Polyacrylamide Gels.

A. Altered Mobility of Heteroduplexes in Native Polyacrylamide Gels.

Amplification reactions (Mullis; Mullis, *et al.*) were performed on peripheral blood mononuclear cell (PBMC) DNA taken from an HIV seropositive asymptomatic man. The reactions were carried out using two sequential 25-35 primer extension cycles, where each cycle used a different set of primers (Example 1). The amplification products were a 1.8 kb (first round) fragment and then a 0.65 kb (second round) internal fragment which corresponded to an internal portion of the HIV-1 envelope gene. When the amplification products were analyzed on an agarose gel, only the expected size band was observed (Example 2, Figure 1B). However, when the amplified DNA was analyzed on a 5% polyacrylamide gel, additional, prominent bands of higher apparent molecular weight were observed (Example 2, Figure 1A).

Given the known variability of HIV-1 within PBMC populations, experiments performed in support of the present invention suggested that the slower migrating bands may have been composed of heteroduplexes formed between divergent molecules during the last melt and reanneal (heat/cool) cycle. To explore this possibility, procedures expected to reduce or eliminate the formation of heteroduplexes were performed. First, the heteroduplexes generated by PCR reactions were subjected to an additional single round of amplification (Example 2, Figure 1, lane 7). This procedure resulted in the loss of the slower migrating DNA bands. These amplification products were remelted and reannealed under conditions that prevented DNA synthesis. In this reaction

mixture, the series of slower migrating bands reappeared (Example 2, Figure 1, lane 8).

Second, the concentrations of the initial PBMC DNA added to the amplification reactions was serially diluted (Example 2, Figure 1, lanes 1-5). This procedure also resulted in the loss of the slower migrating DNA bands. When the highest concentration of DNA was amplified and the resulting products denatured and renatured without amplification, the slower migrating bands were maintained (Figure 1A, lane 6).

In addition, if the additional bands were generated by heteroduplexes, then it seemed reasonable that such heteroduplexes/additional bands would be most efficiently formed at high product concentration, *i.e.*, at later rounds of amplification. As described above, amplification reactions were typically carried out with PBMC DNA as substrate and using two sequential 25-35 primer extension cycles was used to amplify a 1.8 kb and then a 0.65 kb internal fragment of the HIV-1 envelope gene. However, when only 25 cycles of amplification were used in the second round of amplification, the level of slower migrating DNA was less (Example 2, Figure 1, lane 9) than when more cycles of amplification were used. This is result is consistent with the idea that heteroduplex bands are most efficiently formed at high product concentration.

All of the above-results are consistent with the conclusion that the additional bands in non-denaturing polyacrylamide gels were the result of heteroduplex formation during amplification reactions of HIV DNA from infected subjects. The results presented below confirm these observations.

B. The Effects of Gaps and Substitutions on Heteroduplex Mobility.

When two divergent sequences were mixed to form heteroduplex molecules, a single band was observed in agarose gels independent of the degree of sequence-relatedness between the two-strands of the heteroduplex (Figure 2A). However, in polyacrylamide gels, heteroduplexes formed from molecules with mismatched nucleotides, but without gaps, were detected when the degree of divergence reached above 1-2% and generally increased with the degree of mismatch (Figures 2B and 2D). Mixtures of three different sequences yielded six extra bands (Figure 2C). Thus each possible heteroduplex was formed.

Further, heteroduplexes containing single 3 bp internal gaps displayed mobility shifts (Figure 2E). The effect of gapped sequences on heteroduplex mobility was further examined. A number of HIV-1 fragments from different source materials were amplified and fragments having divergent sequences were identified (Example 2, Figure 2E). Based on sequence comparisons, three HIV-1 fragments having internal deletions (*i.e.*, deletions in HIV-1 sequences) of 3 and 9 base pairs were identified (3 and 9, Figure 2E).

Heteroduplexes were formed between the normal fragments, *i.e.*, those not containing a deletion ("insertion fragments", Figure 2E), and the deletion fragments. The size of the gap present in the heteroduplex is shown across the top of the figure. These data show that centrally located gaps of 9 nucleotides resulted in heteroduplexes with slower migration than those with gaps of 3 nucleotides. Further, these results indicate that the mobility of the heteroduplexes is affected by the sequence that is "looped out" of the

"insertion" sequence relative to the "deletion" sequence.

The results presented above demonstrate that the sequence composition of the mismatches and gaps affects the mobility of the heteroduplexes in polyacrylamide gels. This observed effect may be adapted to probe for the presence or absence of specific nucleotides in pre-determined regions of interest, such as activating mutations in proto-oncogenes or particular resistance mutations in viral sequences.

For example, since HMA could detect a 3 bp deletion in a 3.5 Kb DNA fragment it could also potentially be used to localize and utilize polymorphic sites for linkage mapping. Related micro-organisms could also be rapidly classified into sequence homology subgroups using HMA. HIV-1 tissue culture and PCR contamination (problems of frequent occurrence) can be rapidly checked to ensure the identity and purity of the strain under study.

C. <u>The Relationship Between Heteroduplex Mobility Shift and DNA Sequence Distances.</u>

Phylogenetic relationships between genes encoding similar functions are usually described by the DNA distance expressed as the degree of similarity or mismatch between aligned sequences. These measurements ignore unpaired bases or gaps, because of the lack of a suitable method for assigning mutational distance arising from the latter types of mutations (Myers, G., *et al.*, 1985a, 1985b). Having established that heteroduplex mobility shift occurs as a result of both base sequence divergence and gaps, the method of the present invention was used to evaluate the relationship between heteroduplex mobility and DNA

distance measurements based on DNA sequence comparisons.

The method of the present invention was employed to develop simple and rapid assays of the relationships between members of DNA sequence families. In the following examples the heteroduplex mobility assays (HMA) are applied to the study of HIV-1 envelope gene variability, but the method is also applicable to study of other viruses, microorganisms and gene families.

DNA fragments (approximately 700 base pairs) encompassing the V3-V5 region of the HIV-1 *env* gene from a set of 39 molecules of known sequence were amplified (Example 4). These 39 molecules were as follows: thirty one molecules were obtained directly from the PBMC DNA of three North American subjects; six molecules from other independent HIV-1 isolates from North America; and two HIV-1 isolates from Zaire. Heteroduplexes were formed by pairwise combination of sequences from the same individual subject and between epidemiologically unrelated North American and Zairian isolates. The relative mobility of the heteroduplexes compared to the known genetic divergence (Figure 3A). Divergence was determined by standard methods of counting the number of mismatches between aligned sequences, discounting unpaired bases due to insertions and deletions (Kusumi, *et al.*, 1992; Felsenstein, 1988).

Three groupings were observed by both heteroduplex mobility and sequence analysis: (i) large mobility shifts and divergence were observed for heteroduplexes formed between Zairian/North American HIV-1 strains; (ii) intermediate shifts and divergence were observed for heteroduplexes formed between independent North American isolates; and (iii) the smallest shifts and

divergence were observed for comparisons between the closely related sequences derived from the same subjects.

Mobility shifts observed with some sequences derived from a long term infected asymptomatic individual (MA) were atypically large and partially overlapped those observed between independent US isolates (Figure 3A). Further, the mobility shifts were generally lower in the HIV-1 pools from subjects with AIDS than from subject MA. When unpaired nucleotides were counted as mismatches the average genetic divergence increased more for the MA groups than for the AIDS patients. The relatively higher number of gaps distinguishing MA sequences is therefore likely to account for the greater mobility shifts observed. HIV-1 *env* DNA derived from two MA PBMC samples collected 22 months apart were paired to each other for heteroduplex analysis. The mobility shifts of these heteroduplexes increased relative to heteroduplexes formed using the two DNA samples paired to DNA derived from the time point corresponding to when they were isolated. A similar increase in sequence divergence was also observed reflecting the evolution undergone by the virus pool during the sampling interval. These results suggest that the relative levels of HIV-1 pool complexity (genetic divergence plus gaps) can be approximated by pairwise HMA of cloned sequences.

Heteroduplex mobilities were plotted against genetic distance, as determined by percent sequence divergence (Figure 3A). Heteroduplexes differing by only mismatches displayed a better correlation between mobility and divergence but with a lower slope (Figure 3A, circles). The scatter of values in the intra-subject range may

be due to a large influence of gaps on electrophoretic mobility.

In other studies, a more predictable relationship between mobility and DNA distance was observed between heteroduplexes formed between molecules twice as large, encompassing nearly the entire 1.3 kb extracellular envelope coding region.

A standard curve for heteroduplex mobility can be generated using pair-wise combinations of DNA molecules having known genetic distances, based on sequence comparisons between the molecules. For example, a region of a virus, such as the HIV *env* gene, can be selected. A number of variants are sequenced. Heteroduplexes are formed in pair-wise combinations between the variants. The genetic distance (*i.e.*, percent sequence divergence based on mismatch) is then plotted against the heteroduplex mobility for each pair-wise combination. A sample of one such plotted is presented as Figure 8.

Heteroduplexes are then formed between pair-wise combinations of DNA samples, of the same region, by the method of the present invention. Heteroduplex mobilities are determined using HMA and the mobilities compared to the standard curve to establish the degree of relatedness between each pair-wise combination. This information can be used to generate phylogenetic trees such as is shown in Figure 9A.

Figure 9A shows a phylogenetic tree based on HMA. Heteroduplex mobility data was recalculated into genetic distances using a HMA-DNA distance curve (Figure 8). The tree was generated using the "FITCH" program (Phylip 3.4 shareware/freeware) (Felsenstein, 1989). Figure 9B shows a phylogenetic tree based on *env* gene DNA

sequences. Subtypes, identified by Myers (Myers, *et al.*, 1992) are adjacent to boxes. Isolates present in Figure 9A that are not included in Figure 9B are indicated by an asterisk in Figure 9A.

Analyses such as this are useful for the determination of phylogenetic relationships between HIV-1 *env* genes or any other group of related sequences.

II. Applications

The heteroduplex mobility assay (HMA) described herein provides a means for the rapid estimation of the degree of relatedness between members of gene families. HMA requires only a neutral polyacrylamide gel and yields results in hours. The method allows the determination of sequence relatedness without resorting to sequencing analysis (for example, Figure 9A). Under the conditions described above, nucleotide gaps and mismatches can be observed in heteroduplexes using the method of the present invention. The preferred size range for these heteroduplexes is approximately 100 to 1500 base pairs in length, although larger-sized heteroduplexes can be used as well. Heteroduplexes formed from molecules with mismatched nucleotides, but without gaps, were detected when the degree of divergence exceeded approximately 1-2% and generally increased with the degree of mismatch. Typically, heteroduplexes containing an internal gap display mobility shifts which generally increased with the size of the gap. Further, 3 base gaps could be detected in heteroduplexes at least 3 kb in length. Accordingly, the method of the present invention provides the means to determine approximate levels

of DNA sequence diversity in a population of nucleic acid sequences both within and between individuals. The following examples using Human Immunodeficiency Virus sequence variations are illustrative of the method of the present invention.

In addition to obtaining nucleic acid samples by amplification, other samples sources can be used as well. For example, sequences of interest can be cloned (*e.g.*, in a lambda vector; Sambrook, *et al.*) from two different sources. The sequences of interest are independently isolated away from vector sequences (*e.g.*, by restriction endonuclease digestion and fragment purification). These two samples can then be combined, denatured, renatured, and the resulting heteroduplexes analyzed as discussed below.

A. DNA Sequence Divergence Analysis.

1. Estimation of Degree of Sequence Relatedness.

Two previous studies have demonstrated outgrowth of a single genomic variant of *env* with a concomitant reduction in the diversity of HIV genes following the standard virus isolation method of *in vitro* co-culture with uninfected PBMC (Kusumi, *et al.*, 1992).

The method of the present invention was employed to compare the diversity of HIV-1 *env* genes found in PBMC versus those found after co-culture. After co-culture, in 8/12 cases a single genomic variant, a descendant of an originally present variant, was observed. In 2/12 cases two genomic gap-phenotype variants were observed and in 2/12 cases three genomic gap-phenotype variants were observed. In each of the 12 independent cultures a marked reduction of diversity was observed.

When similar analysis was extended to 12 other co-cultures propagated from 2 to 8 weeks and a further 8 un-cultured in vivo quasispecies from different patients the pattern of low sequence complexity following co-culture and higher sequence complexity in vivo was maintained.

In Figure 4 (Example 5), lane "Cult." presents HMA data for DNA samples amplified from a four week co-culture of HIV-infected PBMC. These results show a reduced sequence divergence in HIV-DNA obtained from co-cultured cells relative to the DNA derived from PMBC over time. Further, the comparison between two AIDS patients and an asymptomatic subject (Example 5) demonstrates that the overall level of HIV-1 *env* gene diversity in PBMC DNA (not co-cultured) samples was observed to be lower in subjects with low CD4 cell levels (PE and BU, Figure 4) than in those with higher CD4 cell levels (MA time-points, Figure 4).

The above results demonstrate that *in vitro* PBMC co-culture consistently results in the outgrowth of 1-3 distinguishable variants, confirming two previous studies based on extensive DNA sequence analysis (Meyerhans, *et al.*, 1989; Kusumi, *et al.*, 1992). It is possible that the diverse viral genes observed in HIV-1 infected PBMC represent a record of prior infection in cells that survived an abortive infection with noninfectious virus. Alternatively, the culture milieu may impose selection pressure resulting in preferential replication of viral variants that are able to propagate quickly under the culture conditions. The heteroduplex mobility assay of the present invention may be used to distinguish between these possibilities: HMA can be used to effectively monitor the growth kinetics of specific variants *in vitro*.

The data presented above also demonstrate that *env* gene complexity within the blood of five subjects with AIDS or low CD4 cell numbers was lower than in six asymptomatic seropositive subjects. Outgrowth of a limited number of fast replicating variants in the presence of a declining immune system has been previously implicated in this apparent lower quasispecies complexity (Kuiken, *et al.*, 1992; Tersmette, *et al.*, 1989).

2. Tracking Sequence Variants.

Experiments performed in support of the present invention have led to the development of an assay which rapidly estimates DNA sequence variation without large scale DNA sequencing and which allows tracking and quantitation of variant genotypes in mixtures of such variants. Further, the method of the present invention allows such estimates with greater sensitivity than anything but very large scale DNA sequencing would allow. The heteroduplex mobility assay based on heteroduplex formation of PCR products is faster, simpler, and more informative than the currently available procedures (such as RNase A cleavage mismatch). HMA measures the retardation of electrophoretic mobility of DNA molecules in heteroduplex form relative to (matched) homoduplexes.

The general method of the present invention has been described above. Following is a description of the application of the method of the present invention to the analysis of HIV-1 envelope gene diversity, both within individuals, between individuals and between populations. In addition to analysis of viral genome diversity, such as described below for HIV-1, the method of

the present invention can be applied to the analysis of any number of microorganisms including bacteria, parasites, and other infectious agents. Exemplary microorganisms include, but are not limited to, the following:

(i) Bacterial. *Haemophilus* -- outer membrane proteins, *Staphylococcus*, *Chlamydia* -- outer membrane proteins (Dean, *et al.*), *Enterococcus*, *Mycobacterium (Mycobacterium tuberculosis)*;

(ii) Viral. Feline Leukemia Virus (FeLV), Simian Immunodeficiency Virus (SIV), Human Immunodeficiency Virus (HIV), Hepatitis C Virus (HCV); Human papilloma virus (HPV);

(iii) Fungi. Pneumococcus -- Choline dependent Pneumococcal murein hydrolases; 18S rDNA sequences for human pathogenic fungi including Trichophyton, Histoplasma, blastomyces, coccidioides, Pneumocystis (*Pneumocystis carinii*) and Candida (*Candida albicans*) (Bowman, *et al.*);

(iv) Parasites. Onchocerca (Zimmerman, *et al.*), Babesia spp. (Ellis, *et al.*), Giardia spp. (Weiss, *et al.*), Leishmania spp. (Briones, *et al.*), Trypanosoma spp. (Breniere, *et al.*); and

(v) Mycoplasma. Lyme disease, *Mycoplasma pneumoniae* (Kleemola, *et al.*), using, for example, sequences derived from 16S RNA.

Typically, probes for any target nucleic acid can be selected from a region of the microorganism's genomic material, such as rRNA (for example, as in Weisburg, *et al.*). In this way probes can be identified that will form homoduplexes to identify specific species. Formation of heteroduplexes indicates that the sequences that have diverged from the probe sequence.

The method of the present invention can also be applied to the analysis of any nucleic acid

containing entity, including subcellular organelles such as chloroplasts and mitochondria.

Further, the method of the present invention can also be used in screening methods for the evaluation of therapeutic treatments of any of the above microorganisms. The methods disclosed herein are useful for evaluating, in mixtures of nucleic acids (such as, nucleic acids obtained from tissue samples), the effect over time of a disease treatment, on DNA sequence variation of a nucleic acid target sequence associated with the disease. Therapeutic treatments typically are directed to the resolution, elimination, or relief of a disease state, as, for example, caused by a microorganism/infectious agent.

In one embodiment as applied to HIV infection, the heteroduplex mobility assay can be used to establish a base-line of infection in any selected patient before the onset of treatment. Diversity of the HIV virus can be determined as described below for different cell and tissue samples from the patient. Typically, blood and plasma samples are then serially collected from the subject throughout the therapeutic trial.

Changes during treatment in overall diversity of the HIV population can be monitored by the heteroduplex mobility assay. The turnover of specific variants can be monitored during the course of treatment use labeled tracer DNA (*e.g.*, see below, "(b) Quasispecies Replacement within an Individual). The emergence of new variants can also be detected by using specific HIV probes obtained from samples obtained from different time points in the course of the therapy. Representation of such variants in the population of HIV molecules can also be accomplished user

labeled tracer DNA (*e.g.*, Example 7, Tracking HIV-1 Sequence Variants).

In one embodiment, the method of the present invention is used to evaluate treatment of HIV with 3'-azidothymidine (AZT). The present method allows the evaluation of the effect of treatment on (i) quasispecies diversity, and (ii) the representation of particular variants within sample populations, in particular, variants pre-existing in the host that emerge after the onset of AZT therapy. Further, the method allows the determination of the time-frame in which any AZT-resistant mutants arise during the course of treatment. After the identification of such mutants, the methods of the present invention can be used to assess the presence of such mutants in samples obtained before the mutants were observed (using, for example, the tracer method of Example 7). Further, the present methods can be used to follow the fate of such mutants during the course of further treatment, by, for example, dideoxyinosine (DDI).

Tuberculosis (TB) infections are another example of how the method of the present invention can be used to monitor the effects of a disease treatment. The heteroduplex mobility assay can be used to monitor the presence and diversity of strains of *Mycobacterium tuberculosis* growing within an individual. For example, a 383 bp segment of the gene encoding the 65 kDa mycobacterial surface antigen can be amplified (Ghossein, *et al.*, 1992) from samples obtained from a patient under treatment and analyzed by HMA and/or HTA.

The assay can also be used to detect the specific loss or increase in abundance of TB variants during therapy. This is accomplished by

the tracer probe method of the present invention (Heteroduplex Tracking Assay (HTA); *e.g.*, Example 7) where the labeled probe is derived from standard TB strains or variants identified in the patient during the course of treatment. Such probes are used to track the representation of different *Mycobacterium tuberculosis* populations over time: before, during and following therapy. The present methods can be used to help identify differential strain specificity to anti-tuberculosis drugs.

Generally, the HTA methods of the present invention are used to monitor when variants come and go within the course of any infection and what the impact of any treatment has on the variant populations. Specific loci associated with drug resistance for a particular microorganisms can be used for tracking different populations of a microorganism using the methods of the present invention, where the variant loci are amenable to detection using HTA.

The present assay can be used to evaluate diversity in cell culture systems and animal models as well as patients.

Phylogenetic relationships can be established by the method of the present invention. Phylogenetic analysis can be carried out with almost any selected genomic sequence, such as, glycolytic enzymes (like phosphoglycerate kinase (Vohra, *et al.*)) or rRNA sequences. Phytogenic relationships between plants can be established, using, for example, sequences derived from plastid ribosomal RNA operons (Wolfe, *et al.*).

Use of the method of the present invention to track sequence variants is described below using HIV-1 as an exemplary microorganism.

(a) DNA Heteroduplex Analysis of Envelope Sequence Quasispecies.

Multiple (>50) HIV-1 envelope sequence variants were simultaneously amplified from peripheral blood mononuclear cell (PBMC) DNA by nested polymerase chain reactions (PCR) over the V3 to V5 regions of the *env* genes. The genetic relationship between different DNA quasispecies was determined by radioactively labeling one quasispecies DNA during PCR and reannealing it with an excess of unlabeled DNA from another quasispecies. Other labeling methods, in addition to radioactive labelling, can be employed: for example, labeling with biotin, digoxigenin, or chemiluminescent labels (Tropix, Inc. (Bedford, MA)).

This procedure ensured that most of the probe tracer forms heteroduplexes with the unlabeled target driver DNA. The resulting heteroduplexes were then separated by polyacrylamide electrophoresis. When the probe and the target DNA quasispecies were identical, only fully complementary homoduplexes and moderately retarded radioactive heteroduplexes were seen reflecting the similarity between probe and target DNA, and the sequence heterogeneity within the quasispecies (Example 6, Figure 5).

When quasispecies from unrelated US sero-positives were similarly probed, no homoduplexes were seen and heteroduplexes migrated slower than did heteroduplexes from the same individual. When a quasispecies from a Zimbabwe sero-positive (Africa) was reannealed with US quasispecies, only very slowly migrating heteroduplexes were seen.

The ability to probe one quasispecies with another allows a rapid determination of their degree of sequence similarity. Such information is valuable for epidemiological studies of

transmission risk factors by confirming or ruling out a source of infection (Ou, *et al.*, 1992). For determining the identity and purity of cultured isolates (*e.g.*, HIV-1 strains) or PCR amplification products reactions: for example, if contamination is suspected (Palca, *et al.*).

(b) Quasispecies Replacement within an Individual.

Quasispecies change within an asymptomatic man was monitored by reannealing HIV-1 envelope DNA isolated at different time points spanning 27 months. The progressive disappearance of labeled homoduplexes and the increasing mobility retardation of the labeled heteroduplexes as time increased between the probe and target quasispecies was observed. These results reflect the replacement of one set of sequences by another within two years (Figure 6, Example 6B). The apparent sequence divergence between quasispecies separated by 27 months of *in vivo* evolution was still lower than that observed between epidemiologically unrelated infected individuals from the US.

(c) Tracking Sequence Variants in Complex DNA Mixtures.

Sequential PMBC samples were obtained from an HIV-positive individual. Cloned variant sequences were used as probes to determine the prevalence of the cloned variants in the sequential PBMC samples. In these experiments the presence of radioactive homoduplexes indicated the presence of variant(s) highly related to the probe, *i.e.*, the cloned variants. Cloned variant sequences separated by 22 months of *in vivo* evolution were used to perform heteroduplex probing of six sequentially obtained PBMC samples collected over

27 months (Example 7). Some variants were detected only in the quasispecies from which they were cloned (from independent PCR reactions) (Figure 7A, MA20 and Figure 7C, MA305). Another cloned variant was found in PBMC samples obtained subsequent to the isolation of the cloned variant (Figure 7B, MA16); up to 22 months after the sample from which the cloned variant was isolated.

A variant from the later time point was not detected in the preceding quasispecies (Figure 7C, MA305). The patterns seen using molecular clones as probes, therefore, correspond to those seen using more complex probes (Figure 6, Example 6). However, use of the cloned probes provides a more detailed view of quasispecies change.

The method of the present invention also provides a means to monitor the evolution of sequence complexity in cultured cells as well as *in vivo*. The method of the present invention has been used to track the presence of quasispecies of HIV-1 obtained from PMBC co-cultures. In addition to HIV-1 tracking, the method of the present invention can be applied to variant tracking for other microorganisms.

(d) Evaluation of *in vivo* Quasispecies Complexity.

The method of the present invention is also useful to analyze the *in vivo* evolution of quasispecies complexity for a given microorganism.

PBMC DNA samples from four infected individuals prior to their sero-conversion were analyzed. From two samples, single molecules of HIV-1 DNA were amplified: 5 and 7 positive PCR end point sequences were obtained from these two samples. Pair-wise heteroduplex analysis of these PCR end points, using two different probes, showed no mobility shifts. This result indicates that

these pre-seroconversion quasispecies were made up of identical or very similar sequences.

For two other quasispecies, a higher HIV-1 DNA load allowed the simultaneous amplification of more than 100 envelope sequences pre-seroconversion. In these samples the level of sequence complexity was directly estimated by polyacrylamide gel electrophoresis separation of its homoduplexes and heteroduplexes. This analysis showed these two pre-seroconversion quasispecies to be highly homogeneous.

Similar sequence complexity analysis on sequential PBMC quasispecies was also performed. For two individuals (1058 and 527), the pre-seroconversion HIV-1 DNA load in PBMC fell with the first signs of an antibody response to viral core antigens. Patient 537 HIV-1 DNA developed slight heteroduplex mobility shifts 5.5 months following sero-conversion. Three months later, the quasispecies showed more substantial variation concomitant with an increase in HIV-1 DNA load.

Patient 1058 first showed signs of low level envelope sequence complexity three months after the initial antigenemia peak. Further increase in quasispecies complexity was not apparent until 54 months later, simultaneous with a noticeable increase in HIV-1 DNA load and a detectable p24 antigenemia. The result that HIV-1 load (PBMC HIV-1 DNA) fell with sero-conversion and then within months rose to intermediate levels is consistent with previous observations (Daar, *et al.*; Clark, *et al.*).

Highly homogeneous pre-seroconversion quasispecies DNA was used to probe subsequent quasispecies. Pre-seroconversion sequences were found in 537 PBMC at high levels for 14.5 months post seroconversion and at reduced levels during

the next 32.5 months. At 54 months post-infection the quasispecies appeared free of the pre-seroconversion variant.

Patient 1058, whose quasispecies appeared to remain homogeneous longer, retained sequences highly related to the major pre-seroconversion variant at high levels for the entire observation time of 57 months. A noticeable mobility shift beyond 21 months reflected the accumulation of nucleotide substitutions, which results in small mobility shifts, relative to the major infecting variant and the replacement of the quasispecies with such variants. Sequencing of PCR end points from each patient confirmed that (i) the pre-seroconversion quasispecies were highly related, and (ii) the later sequences had diverged from the pre-seroconversion sequences.

3. Geographical sequence variation.

The method of the present invention also provides a means to track sequence variation between microorganisms isolated from different geographic locations by taking advantage of the gradual increase in mobility shifts observed with increasing sequence divergence between reannealed DNA strands. Analysis of this kind allows the rapid identification of sequence variants within species with, for example, viruses, such as, HIV and influenza.

DNA fragments encoding the V1-V5 of the HIV-1 envelope gene (*env*) from different geographic origins have been amplified and evaluated by the method of the present invention. Using 20 *env* fragments, whose DNA sequences were known, a plot was made of the heteroduplex mobility shift versus genetic distance. From the resulting curve, estimates were made of the genetic distance value

between unknown sequences based on the mobility shifts of their heteroduplexes. The genetic distance for all possible pairs of sequences from within the same country or continent was estimated. The sequences were as follows: ten sequences from Africa, eighteen from the US/Europe, six from Thailand and five from India. All of these sequences were derived from epidemiologically unrelated individuals. The mobilities of heteroduplexes made using HIV-1 DNA found at the same time within uncultured PBMC from six individuals were also measured.

The results of the mobility shift analyses were as follows. Thailand sequences fell into two groups. In the first group the 4 independent isolates from Northern Thailand (Cheng-Mai) formed a very close cluster whose sequence diversity was comparable to that found within individuals. The two sequences from Southern Thailand (Bangkok), while very divergent from those in Cheng-Mai, were themselves more related.

The 5 independent Indian sequences from Bombay were significantly more related than independent sequences found in the US and Europe. African isolates exhibited high sequence divergence, although some Zambian/Zimbabwe pair-wise combinations fell within the lower US/Europe diversity range. All pair-wise heteroduplex combinations tested using sequences within a subgroup resulted in electrophoretic migration typical of that subgroup (*i.e.*, all 6 possible Cheng-Mai heteroduplex pairs showed a distinct grouping), such that no heteroduplex pair behaved anomalously.

4. Phylogenetic Determinations by Heteroduplex Analysis.

To determine phylogenetic relationships between a number of isolates, heteroduplexes consisting of DNA strands from different sources, such as, different geographic, hospital or patient origins, can be analyzed by the method of the present invention. In the results described below, the first round primers were ED3 and ED14, and the second round primers were ED5 and ED12 (Example 1).

Such a phylogenetic analysis using the method of the present invention has been carried out as described above (Figures 8 and 9). Subgroups of related sequences, transcending national and continental barriers, were apparent after mobility shift analysis. For some isolates, distant yet distinct relationships with particular subgroups were observed. Typically, the relatedness between groups was maintained between each pair tested (*e.g.*, GD190, GD132, GD129 with the US/Europe cluster; the Bombay, India with the major Zambia/Zimbabwe clusters).

The subgrouping already apparent from a cursory examination of the mobility shifts was further refined using an algorithm to derive a phylogenetic tree based on heteroduplex mobility analysis (Felsenstein, *et al.*, 1989). All examined US/European isolates clustered in a unique viral subgroup (Figure 9A) together with the two isolates from Bangkok (GD129, GD132) and one from Brazil (GD190). A distinct subgroup was seen for 4 epidemiologically unrelated isolates from Somalia and Zambia and neighboring Zimbabwe (Zim 1, GD178, GD20, GD18-19 [sexual partners]). The sequence cluster from Bombay appeared to be related to this particular Eastern/Southern Africa cluster.

Phylogenetic determinations of HIV-1 isolates from a number of geographical locations showed that the groupings obtained by heteroduplex mobility analysis correspond to those made with actual DNA sequences (Figure 9B). Each of the "GD" HIV-1 isolates studied here and provided under code were grouped in identical clusters by anchored PCR analysis in the *gag* region (samples were obtained from Dr. McCutchan; McCutchan, *et al.*, 1992). It appears that HIV-1 variation was greatest between isolates from the African continent, consistent with a longer residence time for the virus in Africa and larger resulting sequence diversification.

A larger number of other Zambian strains were also reported to cluster using anchored PCR in the gag region (McCutchan, *et al.*, 1992). The Zambia/Zimbabwe cluster were most related to the Bombay, India cluster.

US/European sequences formed a large single related group with related sequences in Bangkok, Thailand and Brazil. A single HIV-1 cluster in the US and Europe could reflect the descendence of most US/European isolates from one or a few highly related sequences in a manner similar to what is now seen in N. Thailand and Bombay.

Independent isolates from Northern Thailand formed a very tight sequence cluster. The actual average nucleotide % substitution within four independent Thailand sequences for the V3 to V5 region was 3.6% ranging from 2.6-4.5%, very close to what was found in four intra-patient quasispecies varying on average 2.3, 2.8, 3.3, and 3.8% with a range of 0.15-8.41%.

Three independent HIV-1 sequences from Bombay, India also showed clustering with an average divergence of 6.1 % ranging from 4.8 to

7%. For contrast, 18 independent US sequence varied on average by 11% ranging from 7.6-13.8%.

The ability to rapidly assign a microorganism (*e.g.*, HIV) to groups of sequence homology should assist in determining the number of major subgroups in different locations. Since no tissue culture, subcloning, or sequencing is required for HMA, sequence variation can be rapidly determined by the method of the present invention. If successful vaccination is dependent upon or improved by a close match with the challenge strain the manufacture of vaccines related to local infectious microorganisms could increase vaccine efficacy. Vaccinated individuals, who nonetheless became infected, could be rapidly analyzed to determine the infecting microorganism subgroup. Such analysis should allow a more rapid appreciation of the effect of sequence variability on vaccine efficacy.

Accordingly, the method of the present invention is useful for tracking quasispecies and sequence variations in samples containing viral nucleic acids. The method of the present invention can also be applied to tracking other pathogen-derived nucleic acids, for example, nucleic acids from bacteria, mycoplasma, protozoans, and parasites. Further, the results presented above demonstrate the usefulness of the method of the present invention for phylogenetic analysis and grouping of isolated microorganisms.

This screening method can be applied to a number of microorganisms, as discussed above. The present method is useful in monitoring changing populations of gene sequences during the course of infection and disease. Further, the method is useful for tracking individual genomic variants and assessing overall levels of diversity, and for

transmission of microorganism variants between individuals (*e.g.*, in the case of HIV-1, maternal-fetal transmission) and within populations (*e.g.*, in the case of HIV-1, in proposed WHO vaccine trial sites).

B. Homoduplex Identification.

Another embodiment of the present invention is the use of specific probes to identify variants based on the formation of homoduplex complexes. For example, sequences corresponding to a particular HIV variant can be cloned and amplified. These cloned sequences are then used as a probe against HIV molecules isolated from a number of test sources. Using the method of the present invention, if homoduplexes are formed in hybridization reactions between the probe and the test source HIV, then the test source HIV is shown to be similar to the cloned probe variant. If on the other hand heteroduplexes are formed between the probe and test sequences, then sequence divergence between the probe and test sequences is indicated.

A battery of probe sequences can be established for major variants and then used in routine screening of test samples to identify the particular variant present in the test sample.

As another example, probe sequences can be selected from the mycoplasma 16S rRNA for a number of species. Probe sequences for each species are generated by polymerase chain reaction using primer sequences conserved between mycoplasma species, but which flank regions of sequence divergence. DNA sequences corresponding to mycoplasma 16S rRNA sequences are available in the "GENBANK" data base for many mycoplasma species. A nucleic acid sample derived from a test source,

such as mycoplasma infected tissue culture, is then amplified with the same primers. The resulting amplified test DNA is separately mixed with tracer amounts of labeled probe DNA (Example 6), denatured, renatured and resolved on polyacrylamide gels. Presence of homoduplexes indicates which species of mycoplasma is present in the infected source.

This method is a self-confirming assay, in that, homoduplex formation should occur with only one probe sequence (representing one species) and heteroduplex formation with the other probes.

C. Gapped Probes.

Gapped probe molecules can be generated to regions of known sequence variation in a target gene. The mobility of heteroduplexes, formed between a target strand and a second strand containing an internal deletion relative to the target strand (*i.e.*, the gapped probe), is affected by the sequence of looped out and the neighboring DNA sequences (Example 3, Figure 2E). Accordingly, the mobility of such heteroduplexes can be used to evaluate the sequence present in the target molecule that bridges the gap.

For example, the use of gapped probe molecules provides a way to identify the presence of mutations in a sample target strand relative to standard target strand that contains a wild-type sequence. This method is particularly useful in the analysis of genes where the location of important mutations has been pre-determined.

D. Oncogene Tracking.

With respect to cancer, once a diagnosis has been made, and a region of DNA associated with the cancerous growth has been identified, the

heteroduplex tracking assay (HTA) of the present invention (*e.g.*, Example 7) can be used to evaluate the extent of infiltration of tumor cells within a tissue population. Exemplary potential target sequences are protooncogenes, for example, including but not limited to the following: c-myc, c-myb, c-fos, c-kit, ras, and BCR/ABL (*e.g.*, Gazdar, *et al.*, 1990; Wickstrom; Zalewski, *et al.*, 1993; Calabretta, *et al.*, 1992, 1993;), oncogenes/tumor suppressor genes (*e.g.*, p53, Bayever, *et al.*). In tumor cells, deletions, insertions, rearrangements and divergent sequences in such genes or in the regions of DNA surrounding the coding sequences of such genes, all allow formation of heteroduplexes between amplified variant DNA and amplified DNA from normal cells.

Specific probes can be designed to the variant oncogenic gene and the labeled probe can be used in HTA. The advantageous sensitivity of HTA is used to detect tumor cells within a population by heteroduplex tracking with sensitivities routinely as low as 1 in 100 cells. Thus, therapies that affect his prevalence can be ascertained through the use of HTA.

The following examples illustrate, but in no way are intended to limit the present invention.

## Materials and Methods

*E. coli* DNA polymerase I (Klenow fragment) was obtained from Boehringer Mannheim Biochemicals (Indianapolis, IN). T4 DNA ligase and T4 DNA polymerase were obtained from New England Biolabs (Beverly, MA); Nitrocellulose filters were obtained from Schleicher and Schuell (Keene, NH). Restriction enzymes were purchased from commercial

vendors and used as per the manufacturer's instructions.

Synthetic oligonucleotide linkers and primers were prepared using commercially available automated oligonucleotide synthesizers. Alternatively, custom designed synthetic oligonucleotides may be purchased, for example, from Synthetic Genetics (San Diego, CA). DNA labeling kits can be obtained from Boehringer-Mannheim Biochemicals (BMB, Indianapolis, IN) and Bethesda Research Laboratories (Gaithersburg, MD). Polymerase chain reaction reagents and equipment are available from Perkin Elmer Corporation (Norwalk, CT).

Routine molecular biology manipulations were carried out by standard procedures as taught, for example, by Ausubel, *et al.*, Maniatis, *et al.*, or Sambrook, *et al.*

## EXAMPLE 1

### Heteroduplex Formation and Analysis

#### A. Cellular DNA Extraction and Polymerase Chain Reactions.

Peripheral blood mononuclear cells (PBMC) isolated by "FICOLL/HYPAQUE" density gradient centrifugation were washed twice with phosphate buffered saline (PBS; Gibco-BRL, Gaithersburg MD). DNA was extracted from the PBMC using the "ISOQUICK DNA" isolation kit (MicroProbe Corp., Garden Grove, CA).

Polymerase chain reactions (PCR) (Mullis; Mullis, *et al.*) were carried out in two rounds using nested primers. Typically, 2 ul of the first round reaction product was added to a second round of PCR with internally annealing primers. First round primers were typically selected from one of the following groups: ED3 (SEQ ID NO:1), corresponding to positions 5956-5985 on the HIV-1-

HXB2 genome (Ratner, *et al.*) and ED12 (SEQ ID NO:2), corresponding to the complement of positions 7810-7781 on the HIV-1-HXB2 genome; or ED3 and ED14 (SEQ ID NO:7), corresponding to positions 7936-7966 on the HIV-1-HXB2 genome.

One set of second round primers was ES7 (SEQ ID NO:3), corresponding to the M13 universal primer followed by positions 7000-7020 of HIV-1-HXB2) and ES8 (SEQ ID NO:4), corresponding to the complement of the M13 reverse sequencing primer followed by positions 7667-7647 of HIV-1-HXB2). The second round primers give rise to an amplification product of 704 bp of which approximately 627 bp are template dependent: the actual size of the amplification products depends on the size of deletions and insertions within the target molecule relative to the HIV-1-HXB2 template.

A second set of second round amplification primers contain the sequences ED5 (SEQ ID NO:8), positions 6562-6588 of HIV-1-HXB2, and ED12 (SEQ ID NO:2), complement of positions 7792-7822 of HIV-1-HXB2. For the geographic analysis, first round primers contained ED3 and ED14 and second round primers contained ED5 and ED12.

A third set of second round amplification primers were ES5 (SEQ ID NO:5), positions 7521-7540 of HIV-1-HXB2, and ED4 (SEQ ID NO:6), complement of positions 7741-7712 of HIV-1-HXB2. Use of primers ES5 and ED4 yielded an ~220 bp amplification fragment.

Each PCR reaction employed variable amounts of template DNA (up to 1 $\mu$g), 1.8 mM MgCl2, 20 pmole of each primer in 50 mM KCl, 10 mM Tris-HCl pH8.3, 200 uM of each dNTP, 2.5 units of Taq DNA polymerase (Perkin Elmer-Cetus, Emeryville, CA), and 10% glycerol in a final volume of 50ul. PCR

reactions were carried out, essentially as per the manufacturer's suggestion, in a Perkin Elmer Thermocycler Model 124 for 25-35 cycles using 1 sec ramp times between steps of 94°C for 1 minute, 57°C for 45 sec, and 72°C for 1 minute.

For heteroduplex resolution a 5 minute 72°C extension step was linked to the last cycle. Heteroduplex resolution was carried out by transferring 10 μl of the second round PCR reaction to 90 μl of fresh standard PCR reaction mix (with 100 pmole of each primer) followed by one denaturation and 5 minute extension cycle at 72°C.

B. End point dilution determination of viral DNA load.

Titration of HIV-1 DNA in PBMC DNA was performed essentially as described by Simmonds, *et al.*, (1990). Briefly, HIV-1 DNA was titrated using duplicate serial 5 fold dilutions of input infected PBMC DNA and maintaining a constant 1 ug of human genomic DNA.

The lowest concentration of infected cell DNA to consistently yield a positive HIV-1 PCR signal was used to estimate the proviral DNA load. When a low proviral DNA load precluded simultaneous amplification of at least 20 template molecules, products from multiple reactions, were pooled. As controls to demonstrate single molecule template sensitivity, pNL4-3 (HIV-1 Lai) DNA (Adachi, *et al.*) and ACH-2 cell DNA (containing a single defective HIV-1 genome) (Folks, *et al.*) were used as infected DNA sources and diluted in the presence of a total of 1 μg of human genomic DNA.

Single molecule HIV DNA templates from subject PBMC were also derived by endpoint dilution and used to generate probe sequences for tracking their representation in PBMC DNA over

time. Endpoint reactions were subjected to the heteroduplex mobility assay of the present invention to verify template homogeneity.

PCR and cloning artifacts were minimized by sequencing HIV-1 variants directly from PCR fragments derived by nested PCR from single molecules of HIV-1 (Simmonds, *et al.*, (1990)). This procedure also provides variant frequencies representative of the prevalence of the HIV-1 molecules within the quasispecies; as may not happen following subcloning, when a single provirus can give rise to two or more subclones.

DNA sequencing of isolated amplification products or clones of HIV DNA was carried out by standard dideoxy sequencing reactions (U.S. Biochemical, Cleveland OH).

C. Heteroduplex formation and analysis.

Typically, 4.5 μl of PCR product from 2 separate reactions were combined and 1-1.5 μl of 10X annealing buffer added (1M NaCl, 100mM Tris HCl pH 7.8, 20 mM EDTA). The mixtures were heated to 94°C for 2-5 minutes and rapidly cooled to 22°C in the thermocycler (1 second ramp). For analysis of HIV samples with high sequence divergence (*e.g.*, samples from different geographical locations) the mixtures were heated to 94°C for 2-5 minutes and rapidly by placing the samples in an ice bath.

Heteroduplexes were then separated on 5% polyacrylamide gels (30:0.8 acrylamide:Bis) at 250V for 3 hours in 1X TBE (0.088M Tris-borate, 0.089M boric acid, 0.002M EDTA) or 2.5% agarose gels at 100V for 1.5 hrs. in TAE buffer (0.04M Tris-acetate, 0.001M EDTA). The gels were stained with ethidium bromide and photographed. The temperature at which the gels are maintained

during electrophoresis can affect the mobility of the heteroduplexes.

EXAMPLE 2

Reduced Mobility of 650 bp DNA Heteroduplexes in Native Polyacrylamide Gels

Nested PCR using two sequential 25-35 primer extension cycles was used to amplify a 1.8 kb and then a 0.65 kb internal fragment of the HIV-1 envelope gene directly from peripheral blood mononuclear cell (PBMC) DNA taken from an HIV seropositive asymptomatic man (Example 1). When the sample was analyzed on a 2.5% agarose gel, only the expected size band was observed (Figure 1B, lanes 1-5). However, when the DNA was analyzed on a 5% polyacrylamide gel (Example 1), additional, prominent bands of higher apparent molecular weight were observed (Figure 1A, lanes 1-5).

To determine the nature of these additional bands, the following analyses were carried out. First, a fraction of the PCR reaction was subjected to a single additional round of PCR using fresh polymerase and an excess of primers (Figure 1, "Resolve", lane 7). Second, the concentration of PBMC DNA was serially reduced in the PCR reactions (Figure 1, lanes 1-5).

Each of these manipulations resulted in the loss of the slower migrating DNA bands (Figure 1A). This result supports the conclusion that the additional bands seen in the initial PCR reactions were heteroduplexes formed between divergent molecules during the last melt and reanneal (heat/cool) cycle of the PCR reactions.

Figure 1A, lane 6 shows the result of melting and reannealing an aliquot of the sample that was fractionated in lane 5. This melting and annealing was not accompanied by an additional

round of amplification. As expected, the number of bands in each of lanes 5 and 6 (Figure 1A) are essentially the same.

Furthermore, the DNA which had been subjected to a single additional round of PCR using fresh polymerase and an excess of primers (sample used in Figure 1, lane 7) was remelted and reannealed in the presence of EDTA to prevent Taq polymerase activity. When this sample was resolved on a 5% polyacrylamide gel, the series of slower migrating bands reappeared (Figure 1A, Heat/Cool, lane 8).

In addition, nested PCR was carried out using only one round of 25 primer extension cycles to amplify a 0.65 kb internal fragment of the HIV-1 envelope gene directly from the PBMC DNA-described above. In these reactions, the products demonstrate that the level of slower migrating DNA, *i.e.*, the additional bands, was reduced (Figure 1, lane 9) relative to the 35 cycle amplification described above.

Each of the results presented above are consistent with heteroduplex formation during later stages of amplification when HIV DNA from an infected subject is used as template DNA.

## EXAMPLE 3

### Heteroduplexes Containing Sequence Gaps or Base-Pair Mismatches Demonstrate a Mobility Shift in Polyacrylamide Gels

A. Mismatch and Gaps.

Heteroduplexes were formed by melting and reannealing mixtures of DNA fragments essentially as described above. Samples of HIV-1 were obtained from a variety of sources. Briefly, molecular clones of HIV-1 genes in plasmids were obtained (*e.g.*, Kusumi, *et al.*). Sequences were either obtained from published sequences or by employing standard sequencing techniques with

using isolated HIV sequences. Heteroduplexes were formed as described in Example 1. The migration of the heteroduplexes and homoduplexes were compared in polyacrylamide and agarose gels as described in Examples 1C and 2.

Figures 2B and 2C show photographs of ethidium bromide stained 5% polyacrylamide gels containing, in each lane, heteroduplexes formed from a mixture of two PCR products, each PCR amplified product has been sequenced. The level of sequence diversity between the two products is indicated in the figure: sequence percent mismatch ranges from 0.16 to 1.11. Also, the presence (+) or absence (-) of gaps in the heteroduplex sequences are indicated.

In Figure 2D results are shown for heteroduplexes having the levels of sequence diversity indicated in the figure: percent mismatch from 1.3 to 4.9.

When two divergent sequences were mixed, a single band was observed in agarose gels regardless of the sequence relationships between the sequences (Figure 2A). In polyacrylamide gels the same mixtures resulted in nearly comigrating homoduplex bands (bottom band in each lane, Figures 2B and 2C) plus two additional slower migrating bands (Figures 2B and 2C).

Mixing and annealing three different-sequence amplification products yielded six bands (Figure 2D, lane 1) in addition to the homoduplex band (which corresponds to the fastest migrating band in Figure 2D, lane 1). Thus each possible heteroduplex is formed. Figure 2D, lane 2 contains molecular weight markers derived from a *HaeIII* digest of $\phi$X174.

The results presented above suggest that the composition of the mismatches and gaps affects

mobility. The effect of different sized gaps on heteroduplex mobility was next examined.

B. The Effects of Sequence Variations Within Insertion Sequences.

A number of HIV-1 fragments from different source materials were amplified (second round primers ES7 and ES8) and sequenced as described above (Example 1). The amplification products were normally 704 bp of which approximately 627 bp are template dependent. A number of 704 bp fragments having divergent sequences were identified ("Insertion" fragments, A-H, Figure 2E). The duplicate lanes correspond to fragments having the same nucleotide sequence that were derived from different sources. Further, based on sequence comparisons, three HIV-1 fragments having internal deletions (*i.e.*, deletions in HIV-1 sequences) of 3 and 9 base pairs were identified (3 and 9, Figure 2E), which were designated MA21 (3 base pairs (bp)), MA311 (9 bp), and MA6 (9 bp).

Heteroduplexes were formed between the insertion fragments and the deletion fragments. The heteroduplexes were electrophoretically separated as described in Example 1. The gels were stained with ethidium bromide and photographed. The results are shown in Figure 2E. In the figure, the bottom band in each lane corresponds to homoduplexes. The lanes marked "M" are molecular weight standards.

The size of the gap present in the heteroduplex is shown across the top of the figure. These data show that centrally located gaps of 9 nucleotides resulted in heteroduplexes with slower migration than those with gaps of 3 nucleotides. Further, these results indicate that the mobility of the heteroduplexes is affected by the sequence that is "looped out" of the

"insertion" sequence relative to the "deletion" sequence. For example, heteroduplexes formed with deletion fragment 12 have a range of mobilities where the mobilities are dependent on the sequences present in the looped out insertion of the second fragment A, B, C or D.

EXAMPLE 4

Heteroduplex Mobility Shift and DNA Sequence Distances

A number of HIV-1 fragments, corresponding to HIV-1 *env* sequences (V3-V5) from within the same and between different seropositive individuals, were amplified (second round primers ES7 and ES8) and sequenced as described above (Example 1). The amplification products were normally 704 bp of which approximately 627 bp are template dependent.

Relative mobilities of HIV-1 DNA intra- and inter-subject heteroduplexes, formed using amplified DNA from the following sources, were evaluated by the method of the present invention: MA89-91 is a comparison of sequences derived from subject MA 22 months apart in 1989 and 1991; US/US is a comparison of sequences from epidemiologically distinct viruses from subjects within the United States of America; US/AFR is a comparison of sequences from 6 US [pNL4-3, SF2, SF162, MA5 (Myers, *et al.*, 1992); BU01, PE01] and 2 Zairian (Africa) subjects [NDK, MAL (Myers, *et al.*, 1992)]. Included in the US/AFR group is the NDK/MAL comparison which displayed the fastest mobility and least sequence divergence in the group.

For each comparison of heteroduplexes formed using amplification products from two source DNAs, heteroduplex mobilities on non-denaturing polyacrylamide gels were calculated as the average distance of migration of the two heteroduplex

bands divided by the distance of migration of the homoduplex bands. The sequence divergence between the nucleic acid sequences from each HIV-fragment source were determined by comparison of the nucleic acid sequences using the program DOTS (Kusumi, *et al.*, 1992). The program counts the number of mismatched bases between aligned sequences, discounting gaps introduced to maintain alignment.

In Figure 3A the relative mobility values are plotted against the percent divergence for four sets of sequence comparisons: +, Intrasubject, from within subjects MA, PE and BU; •, Intrasubject, including only comparisons for which no unpaired segments or gaps appear within heteroduplexes; □, US/US inter-subject; ◇, US/AFR inter-subject.

Figure 3B shows representative heteroduplex mobility data using the above DNA sources and a 5% polyacrylamide gel with representative heteroduplexes formed from the intrasubject, intersubject and US/AFR groups, respectively.

The above data demonstrate the correlation between the degree of sequence relatedness, based on direct nucleic acid sequence comparisons, and heteroduplex/homoduplex mobility on non-denaturing polyacrylamide gels: generally, the relative mobility shift increases with increasing percent divergence within patient samples.

## EXAMPLE 5

### Consistent Reduction of Pool Diversity Upon PBMC-Co-Culture

The heteroduplex mobility assay described above was used to compare the diversity of HIV-1 *env* genes found in isolated PBMC versus those found after co-culture of the PMBC (Kusumi, *et al.*, 1992).

DNA was isolated from PMBC as described in Example 1. Titration of the HIV-1 DNA in the PBMC DNA was performed as described in Example 1. Each PBMC DNA sample contained at least 20 molecules of HIV-1 DNA, determined by end point PCR. In each sample the HIV-1 molecules were simultaneously PCR amplified as described above using ES7 and ES8 as second round primers in nested PCR reactions.

Figure 4 shows the results of heteroduplex mobility analysis of viral quasispecies obtained from PBMC and after virus isolation in co-culture. In the figure, lane numbers refer to the month of collection of PBMC from an asymptomatic subject (MA) beginning approximately 5 years after infection (Kusumi, *et al.*, 1992). Duplicate nested PCR reactions are shown for the 1 and 7 month time points. "Cult." refers to DNA samples amplified from a four week co-culture of HIV-infected PBMC (time point, 1 month) with uninfected PBMC. Lane "M" contains molecular weight markers. "PE" and "BU" correspond to source PMBC DNA obtained from two AIDS patients.

The results presented in Figure 4 indicate a reduced sequence divergence in HIV-DNA obtained from co-cultured cells relative to the DNA derived from PMBC over time. Further, the PMBC DNA samples from the two AIDs patients, PE and BU, also demonstrate a lower sequence divergence relative to the amplified samples from the asymptomatic patient.

## EXAMPLE 6

### Radioactive Probes and HMA

#### A. Relationships between Quasispecies from Epidemiologically unlinked Individuals.

Typically, multiple (>50) HIV-1 envelope sequence variants were simultaneously amplified (Example 1) from peripheral blood mononuclear cell

(PBMC) DNA by nested polymerase chain reactions (PCR) over the V3 to V5 regions of the *env* genes (Example 1, using ES7 and ES8 as second round primers in nested PCR reactions).

The genetic relationship between DNA molecules amplified from different source DNAs was determined by radioactively labeling one group of DNA molecules during PCR and reannealing those labeled molecules with an 100-fold excess of unlabeled DNA from another source material.

With this procedure, most or all of the probe "tracer" (*i.e.*, labelled amplified DNA molecules) formed heteroduplexes with the unlabeled target "driver" DNA. The resulting heteroduplexes were then separated by polyacrylamide electrophoresis (Example 1). The gels were then exposed to X-ray film (with or without intensifying screens) and the resulting autoradiograms analyzed.

Figure 5 shows the results of one such "tracer"/"driver" analysis. PMBC DNA was isolated from four sources, three US HIV-positive samples and one African HIV-positive sample (from Zimbabwe). Nested DNA amplification reactions were carried out as described in Example 1 using primers ES7 and ES8.

The probes were labeled PCR products from each PMBC sample. The products were radiolabeled by addition of 10 $\mu$Ci of a $^{32}$P-dCTP and 30 $\mu$M of each dNTP to the second round of a nested PCR reaction.

Unlabeled PCR products from the subject PBMC samples were mixed, heated and reannealed in 100 fold excess with each of the radiolabeled probes. The resulting reannealed products were electrophoretically separated on a 5% polyacrylamide gel. An autoradiograph of four such analyses is shown in Figure 5. The asterisk

(*) to the left of the panel denotes the position of single stranded DNA, which is of variable intensity and pattern from experiment to experiment.

When the probe and the target DNA were the same (Figure 5, lanes 1-1, 2-2, 3-3, and 4-4), only fully complementary homoduplexes and moderately retarded radioactive heteroduplexes (representing the members of this quasispecies) were seen. This result reflects the similarity between probe and target DNA, and the limited sequence heterogeneity within the quasispecies.

When quasispecies from unrelated US sero-positives were similarly probed, no homoduplexes were seen and heteroduplexes migrated slower than did heteroduplexes from the same individual (Figure 5, panels 1-3, lanes 1-3). When a quasispecies from a Zimbabwe sero-positive (Africa) was reannealed with the US quasispecies, only very slowly migrating heteroduplexes were seen (Figure 5, panels 1-3, lane 4, and panel 4, lanes 1-3).

B. Intra-Individual HMA for HIV-1 Quasispecies.

Quasispecies change within an asymptomatic HIV-positive man (MA) was monitored as follows. HIV-1 envelope DNA was isolated by nested PCR amplification, using ES7 and ES8 as second round primers, from PMBC DNA. The PBMC DNA was isolated from the subject at different time points spanning 27 months. "Tracer"/"driver" analysis was carried out as described in Example 5 using amplified DNA from each PMBC time point.

In this example, the source of the PMBC DNA was asymptomatic patient MA, described above in Example 5.

Probes were derived from PCR products from the month 1 (M1, Figure 5, left panel) or month 27 (M27, Figure 5, right panel). Month 1 and 27 products were radiolabeled by addition of 10 $\mu$Ci of a $^{32}$P-dCTP and 30 $\mu$M of each dNTP to the second round of the nested PCR reaction.

Polymerase chain reaction products from the asymptomatic subject MA PBMC samples, described above in Example 5, were mixed, heated and reannealed in 100 fold excess with two radiolabeled probes.

The resulting reannealed duplex products were electrophoretically separated on a 5% polyacrylamide gel. An autoradiograph of the gel is shown in Figure 6. The asterisk (*) to the left of the panel denotes the position of single stranded DNA.

The labeled probe DNA (tracer) is shown at the top of the figure: M1 corresponds to the quasispecies amplified from the first month PMBC and M27 corresponds to the quasispecies amplified from the 27 month PMBC. The driver DNA is shown in the second line of the figure: the number represents the month the sample was obtained.

The progressive disappearance of labeled homoduplexes (M1-tracer) and the increasing mobility retardation of the labeled heteroduplexes (M1-tracer) as time increased between the probe and target quasispecies was observed. The M27-tracer demonstrates the replacement of one set of sequences (M27-tracer:M1-driver) by another (M27-tracer:M27-driver) within two years.

## EXAMPLE 7

### Tracking HIV-1 Sequence Variants

PMBC DNA was isolated from an HIV-positive individual over a twenty seven month period.

Variant sequences were cloned from the 1 month PMBC sample (clones MA20 and MA16) and from the 22 month sample (clone MA305): the source DNAs for these clones were the unamplified PMBC DNA samples. HIV sequences (for example, *env* sequences) were cloned using standard vectors and techniques (Kusumi, *et al.*, 1992; Ausubel, *et al.*; Sambrook, *et al.*). The cloned sequences were labeled with radioactive moieties by end-labeling, random-priming or nick-translation (kits for each method available from Gibco/BRL, Gaithersburg, MD).

The HIV *env* sequences were amplified as described in Example 1 using primers ES7 and ES8 and the sequentially obtained PBMC DNAs as template (as in Example 6). The cloned DNAs were used as probes to determine their prevalence in sequential PBMC samples as described in Example 6. The results of this analysis are shown in Figures 7A, 7B, and 7C.

In Figure 7, the numbers at the top of the lanes correspond to the month that the sequential PMBC sample was isolated. The clone names are indicated at the top of each panel of the figure (Figure 7A is MA20, Figure 7B is MA16, and Figure 7C is MA305). M1 indicates that the probe was obtained from the 1 month PMBC DNA sample. M22 indicates that the probe was obtained from the 22 month PMBC DNA sample. "s.s." is the location of the labeled single-strand DNA and "H" is the location of the homoduplex.

One clone (MA20) obtained from the 1 month PMBC DNA detected the presence of its corresponding variant only in the quasispecies (*i.e.*, the family of variant HIV sequences present in the PMBC sample) from which it was cloned ("H", lane 1, Figure 7A). Another clone, MA305, which

represents a different quasispecies, was present in the amplified PMBC DNA from which it was obtained (Figure 7C, lane 22) and also in two subsequent samples (Figure 7C, lanes 23 and 27).

The third clone, MA16, was found in the amplified PMBC DNA from which it was obtained (Figure 7B, lane 1) and also in subsequent PBMC samples up to 22 months later (Figure 7B, lanes 6, 7 and 22).

In Figures 10B and 10C, the lane marked "C" is a control reaction showing the results of mixing the tracer clone DNA with the driver DNA corresponding to the PMBC sample from which the clone was obtained.

While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: The Board of Trustees of the Leland Stanford Junior University
(C) CITY: Stanford
(D) STATE: CA
(E) COUNTRY: USA
(F) POSTAL CODE: 94305

(i) APPLICANT:
(A) NAME: The Government of the United States of America as represented by The Secretary of the Department of Health and Human Services and His Successors
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) POSTAL CODE: 20231

(ii) TITLE OF INVENTION: A Heteroduplex Mobility Assay for the Analysis of Nucleic Acid Sequence Diversity

(iii) NUMBER OF SEQUENCES: 8

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Dehlinger and Associates
(B) STREET: 350 Cambridge Avenue, Suite 250
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94306

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/241,373
(B) FILING DATE: 11-MAY-1994

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/087,010
(B) FILING DATE: 1-JUL-1993

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Fabian, Gary R.
(B) REGISTRATION NUMBER: 33,875
(C) REFERENCE/DOCKET NUMBER: 8600-0130.41

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 324-0880
(B) TELEFAX: (415) 324-0960

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: PRIMER ED3, HIV-1-HXB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTAGGCATCT CCTATGGCAG GAAGAAGCGG 30

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: PRIMER ED12, HIV-1-HXB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AGTGCTTCCT GCTGCTCCCA AGAACCCAAG 30

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: PRIMER ES7, M13/HIV-1-HXB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TGTAAAACGA CGGCCAGTCT GTTAAATGGC AGTCTAGC 38

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: PRIMER ES8, M13/HIV-1-HXB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CAGGAAACAG CTATGACCCA CTTCTCCAAT TGTCCCTCA 39

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: PRIMER ES5, HIV-1-HXB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CAATGTATGC CCCTCCCATC 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: PRIMER ED4, HIV-1-HXB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CACCACTCTT CTCTTTGCCT TGGTGGGTGC 30

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: PRIMER ED14, HIV-1-HXB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCTTGCCTGG AGCTGCTTGA TGCCCCAGAC 30

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: PRIMER ED5, HIV-1-HXB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGGGATCAA AGCCTAAAGC CATGTG 26

IT IS CLAIMED:

1. A method of evaluating sequence diversity in a mixture of nucleic acids containing a target sequence, comprising

selecting amplification primers complementary to nucleic acid sequences flanking the target sequence,

combining the nucleic acids and the primers under conditions that promote the hybridization of the primers to the nucleic acids, thus generating primer/nucleic acid complexes,

converting the primer/nucleic acid complexes to double-strand fragments in the presence of a suitable polymerase and all four deoxyribonucleotides,

amplifying the number of primer-containing fragments by successively repeating the steps of (i) denaturing the double-strand fragments to produce single-strand fragments, (ii) hybridizing the single strands with the primers to form strand/primer complexes, (iii) generating double-strand fragments from the strand/primer complexes in the presence of DNA polymerase and all four deoxyribonucleotides, and (iv) repeating steps (i) to (iii) until a desired degree of amplification has been achieved,

denaturing and renaturing the amplified fragments to form a population of amplified fragment DNA duplexes,

separating the duplexes on polyacrylamide gels, and

analyzing the relative migration of the duplexes to establish the relative degree of sequence relatedness in the population of amplified fragments.

2. The method of claim 1, where said nucleic acid is RNA and where in said converting the suitable polymerase is reverse transcriptase.

3. The method of claim 1, where said denaturing is thermal denaturing, and said generating is carried out using a thermostable DNA polymerase.

4. The method of claim 1, where said separating is carried out by polyacrylamide gel electrophoresis, and said analyzing involves visualization of the amplification products by ethidium bromide staining.

5. The method of claim 1, where said separating is carried out by polyacrylamide gel electrophoresis, and said analyzing includes transfer of the nucleic acids from the gel to a support membrane, and hybridization of the nucleic acid transferred to the membrane with a labelled probe specific for the desired amplification products.

6. The method of claim 1, where the primers contain at least one detection moiety.

7. The method of claim 6, where said detection moiety is a radioactive moiety or biotin.

8. The method of claim 1, where said nucleic acid containing a target sequence is a nucleic acid from a microorganism.

9. The method of claim 8, where said microorganism is Human Immunodeficiency Virus 1.

10. The method of claim 8, where said method is used to evaluate sequence diversity over time in mixtures of nucleic acids containing a target sequence, where the mixtures are serially obtained from a single source.

11. The method of claim 10, where said microorganism is Human Immunodeficiency Virus 1 (HIV-1), and where said single source is a patient infected with HIV-1.

12. The method of claim 1, where said nucleic acid containing a target sequence is a nucleic acid from an oncogene.

13. A method of any of claims 1 to 12, for evaluating, in mixtures of nucleic acids, the effect over time of a disease treatment, on DNA sequence variation of a nucleic acid target sequence associated with the disease,

where mixtures of nucleic acids are serially obtained from a single source,

and where the method further includes,

evaluating the effect of the treatment by comparing the relative degree of sequence relatedness of amplified fragments in each serial sample.

14. A method of evaluating sequence diversity between two different sample mixtures of nucleic acids, where said nucleic acids contain a target sequence, comprising

selecting amplification primers complementary to nucleic acid sequences flanking the target sequence,

combining each nucleic acid sample individually with the primers under conditions

that promote the hybridization of the primers to the nucleic acid, thus generating primer/nucleic acid complexes,

converting the primer/nucleic acid complexes in each sample to double-strand fragments in the presence of a suitable polymerase and all four deoxyribonucleotides,

amplifying the number of primer-containing fragments by successively repeating the steps of (i) denaturing the double-strand fragments to produce single-strand fragments, (ii) hybridizing the single strands with the primers to form strand/primer complexes, (iii) generating double-strand fragments from the strand/primer complexes in the presence of DNA polymerase and all four deoxyribonucleotides, and (iv) repeating steps (i) to (iii) until a desired degree of amplification has been achieved,

mixing together the amplified fragments from each sample,

denaturing and renaturing the amplified fragments to form a population of amplified fragment DNA duplexes,

separating the duplexes on polyacrylamide gels, and

analyzing the relative migration of the duplexes to establish the relative degree of sequence relatedness among the amplified fragments of the population.

15. The method of claim 14, where said nucleic acid containing a target sequence is a nucleic acid from a microorganism.

16. The method of claim 15, where said the samples of said nucleic acid are collected from a number of different geographic locations.

17. The method of claim 15, where said microorganism is Human Immunodeficiency Virus 1.

18. The method of claim 14, where the amplified fragments of one sample nucleic acid are labelled with a detection moiety and where said labeled fragments are mixed with a molar excess of the amplified fragments of the other sample.

19. A method for detecting the presence of a selected nucleic acid target sequence in a nucleic acid sample, comprising

selecting (i) a duplex DNA probe having two complementary strands, where the duplex is homologous to the target sequence and each strand contains a detection moiety, and (ii) amplification primers complementary to nucleic acid sequences flanking the target sequence of the nucleic acid,

combining the nucleic acid sample with the primers under conditions that promote the hybridization of the primers to the nucleic acid, thus generating primer/nucleic acid complexes,

converting the primer/nucleic acid complexes in to double-strand fragments in the presence of a suitable polymerase and all four deoxyribonucleotides,

amplifying the number of primer-containing fragments by successively repeating the steps of (i) denaturing the double-strand fragments to produce single-strand fragments, (ii) hybridizing the single strands with the primers to form strand/primer complexes, (iii) generating double-strand fragments from the strand/primer complexes in the presence of DNA polymerase and all four deoxyribonucleotides, and (iv) repeating steps (i)

to (iii) until a desired degree of amplification has been achieved,

mixing the duplex probe with a molar excess of the amplified fragments, and

denaturing and renaturing the mix of probe and amplified fragments to form a population of DNA duplexes,

separating the population of DNA duplexes and the duplex DNA probe on a polyacrylamide gel, and

analyzing the migrations of duplexes, which contain a strand of the probe, relative to the migration of the probe duplex, to establish the relative degree of sequence relatedness between the probe and sample target sequences.

20. The method of claim 19, where said mixing is carried out at a ratio of 100:1, amplified fragments to probe.

21. The method of claim 19, where said detection moiety is biotin or a radioactive moiety.

22. The method of claim 19, where said nucleic acid containing a target sequence is a nucleic acid from a microorganism.

23. The method of claim 22, where said microorganism is a Human Immunodeficiency Virus 1.

24. The method of claim 23, where said probe is selected from the genome of Human Immunodeficiency Virus 1.

25. The method of claim 19, where said nucleic acid containing a target sequence is a nucleic acid from an oncogene.

26. A method of any of claims 19 to 25, for evaluating the effect of a disease treatment procedure on the presence of a selected nucleic acid target sequence in a nucleic acid sample,

where samples are obtained from a single source and a first sample is obtained before treatment and a second sample is obtained after treatment,

and where the method further includes,

evaluating the effect of a disease treatment procedure on the presence of a selected nucleic acid target sequence by comparing the relative degree of sequence relatedness between the probe and target sequences of the second nucleic acid sample relative to the first nucleic acid sample.

Fig. 1A

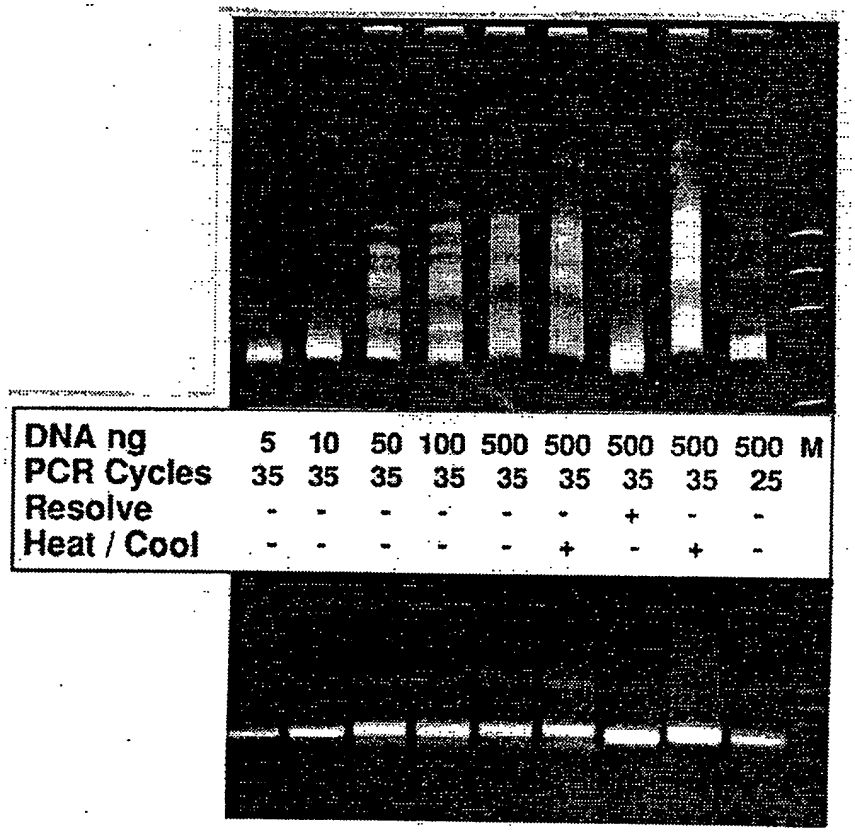

DNA ng 5 10 50 100 500 500 500 500 500 M
PCR Cycles 35 35 35 35 35 35 35 35 25
Resolve - - - - - - + - -
Heat / Cool - - - - - + - + -

Fig. 1B

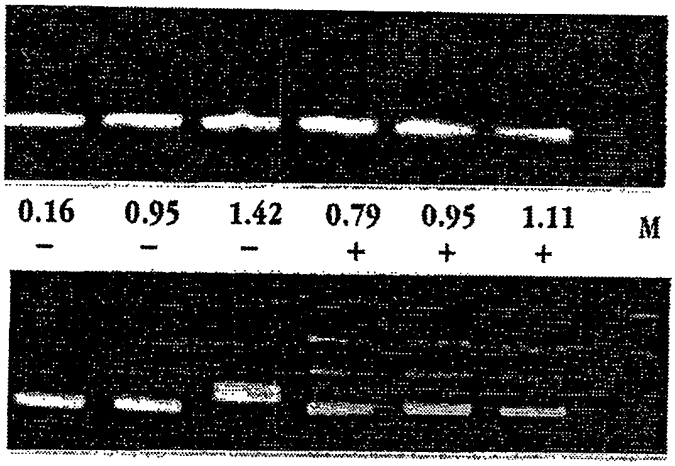



Fig. 2A

Fig. 2B

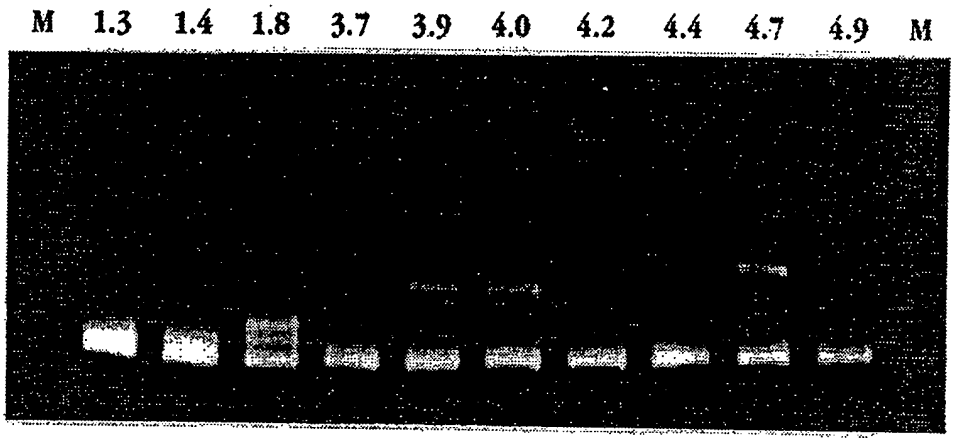



Fig. 2C

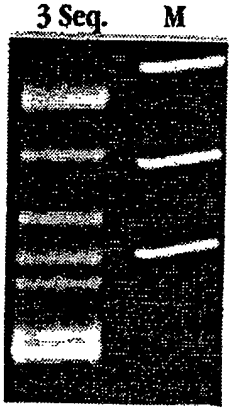



Fig. 2D

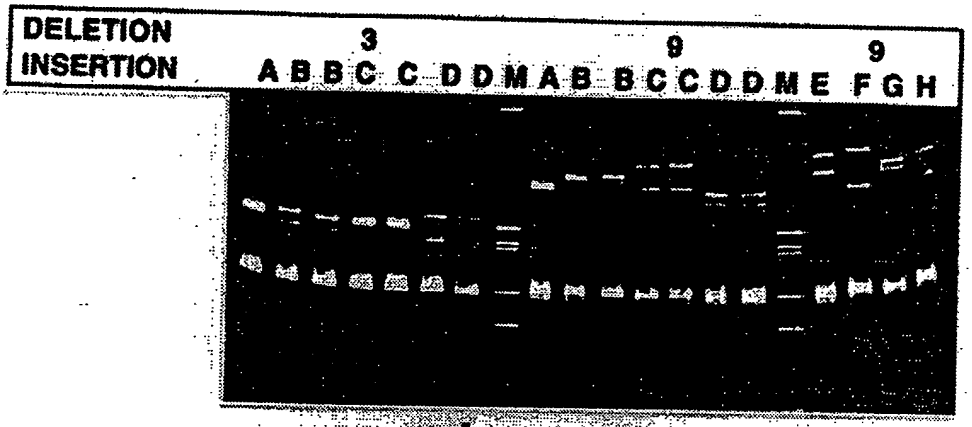

DELETION
INSERTION
3
9
9
A B B C C D D M A B B C C D D M E F G H

Fig. 2E

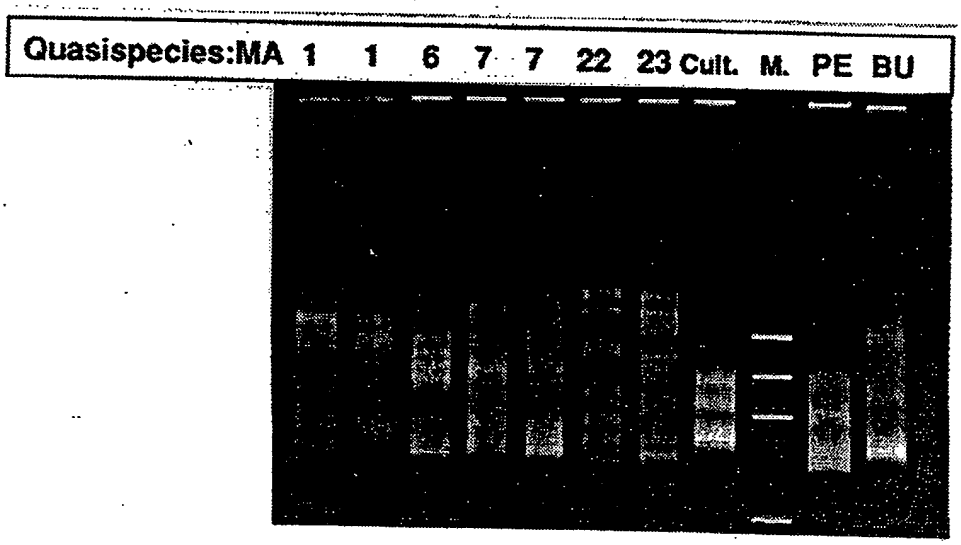

Quasispecies:MA 1 1 6 7 7 22 23 Cult. M. PE BU

Fig. 4

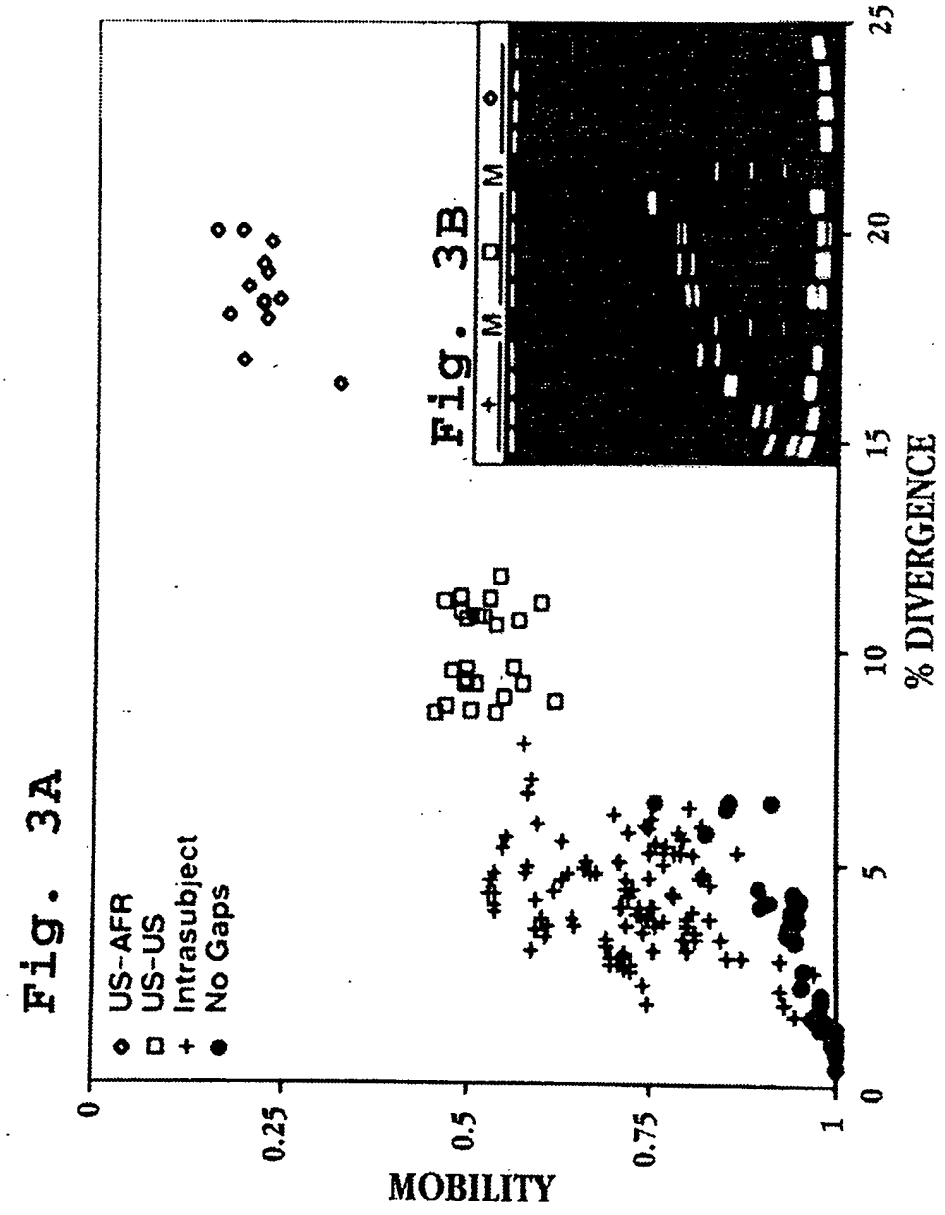


Fig. 3A
US-AFR
US-US
Intrasubject
No Gaps
0
0.25
0.5
0.75
1
MOBILITY
0
5
10
15
20
25
% DIVERGENCE
Fig. 3B
M
M

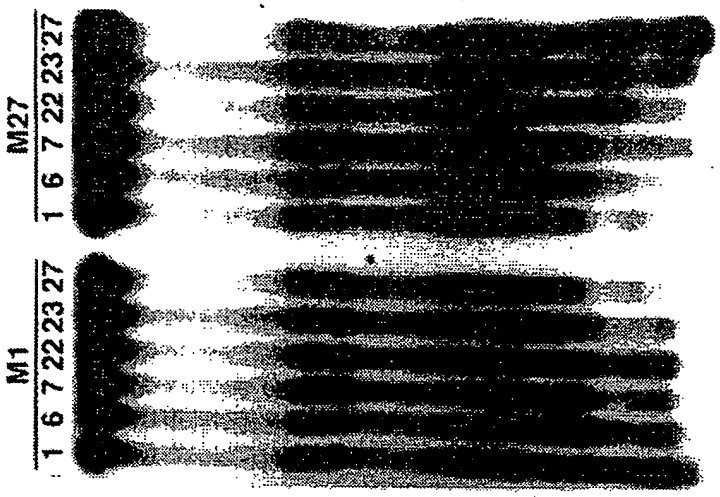



Fig. 6

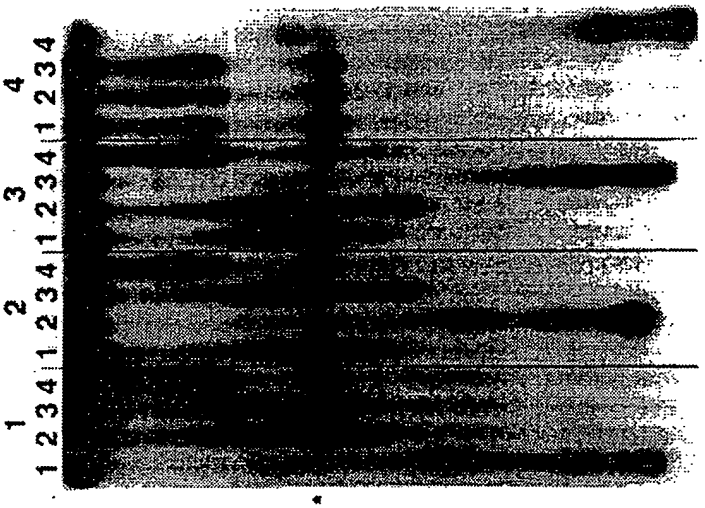



Fig. 5

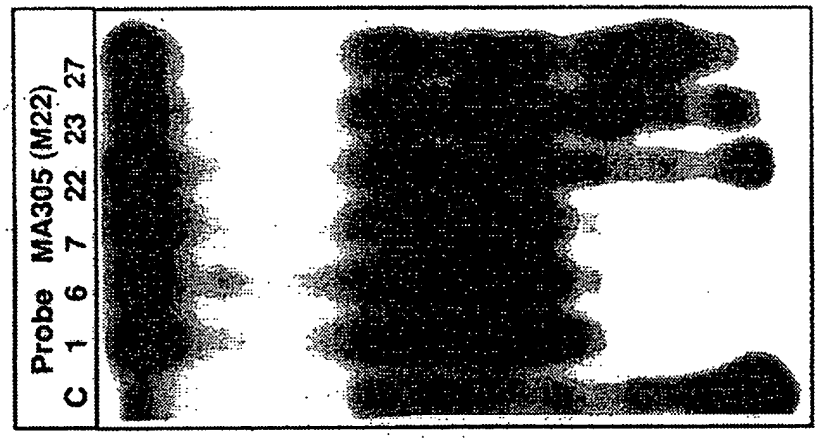



Fig. 7C

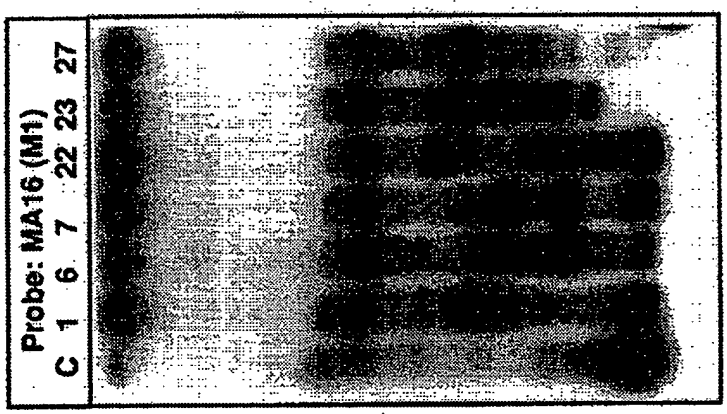



Fig. 7B

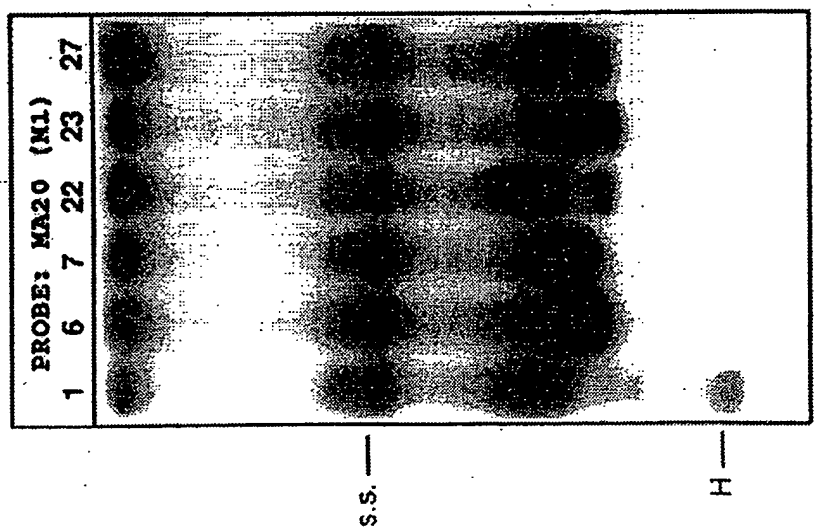



Fig. 7A

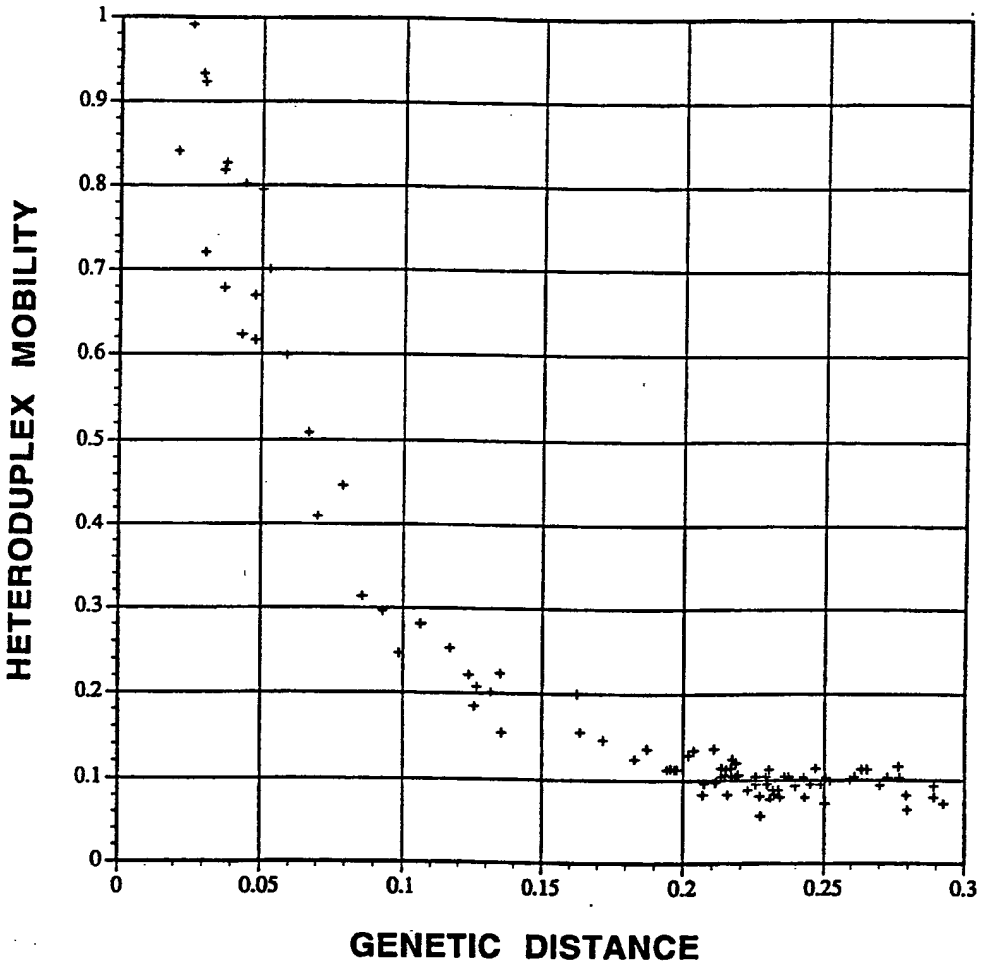

HETERODUPLEX MOBILITY
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.05
0.1
0.15
0.2
0.25
0.3
GENETIC DISTANCE

Fig. 8

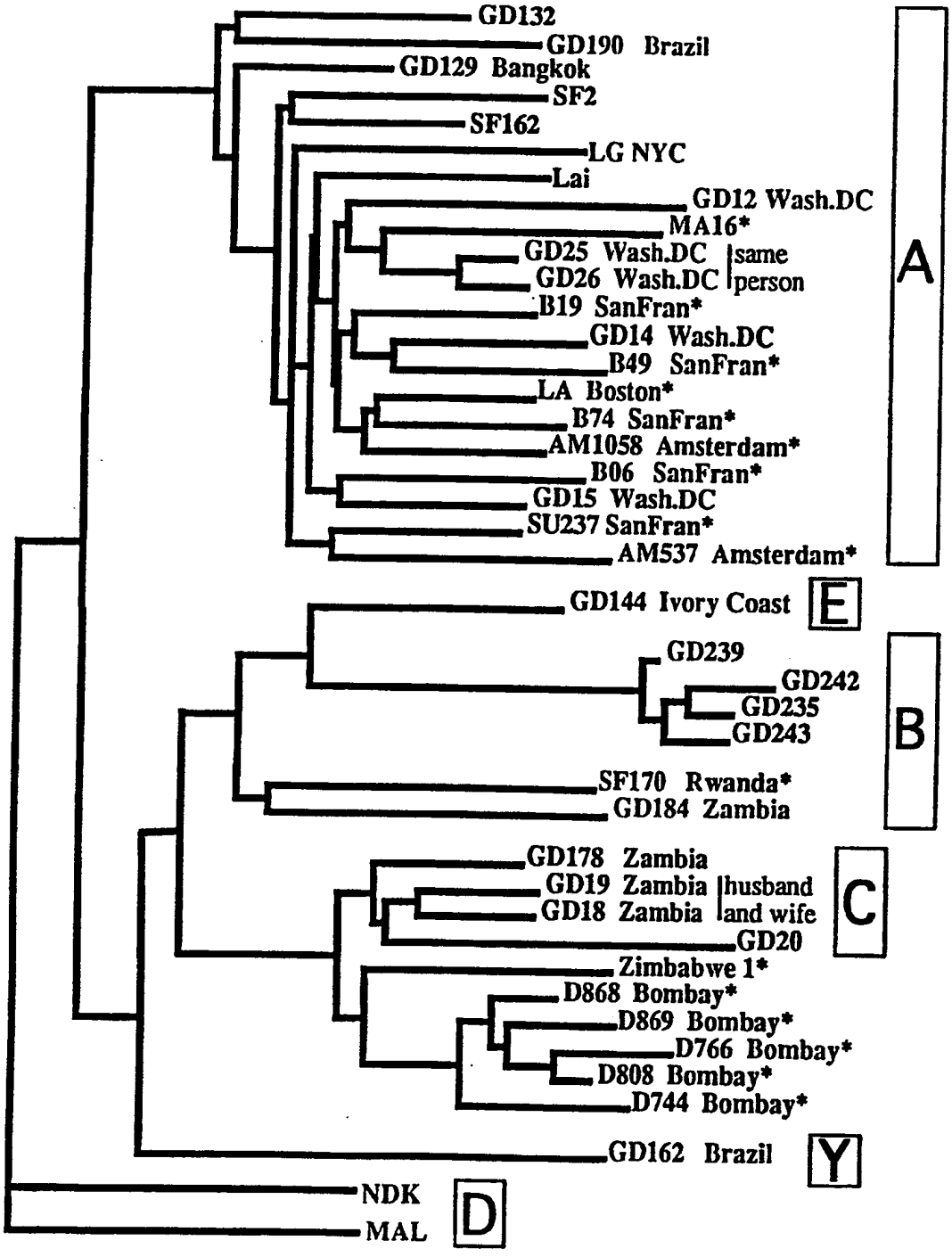

GD132
GD190 Brazil
GD129 Bangkok
SF2
SF162
LG NYC
Lai
GD12 Wash.DC
MA16*
GD25 Wash.DC
GD26 Wash.DC
same person
B19 SanFran*
GD14 Wash.DC
B49 SanFran*
LA Boston*
B74 SanFran*
AM1058 Amsterdam*
B06 SanFran*
GD15 Wash.DC
SU237 SanFran*
AM537 Amsterdam*
A
GD144 Ivory Coast
E
GD239
GD242
GD235
GD243
B
SF170 Rwanda*
GD184 Zambia
GD178 Zambia
GD19 Zambia
GD18 Zambia
husband and wife
GD20
C
Zimbabwe 1*
D868 Bombay*
D869 Bombay*
D766 Bombay*
D808 Bombay*
D744 Bombay*
GD162 Brazil
Y
NDK
MAL
D

Fig. 9A

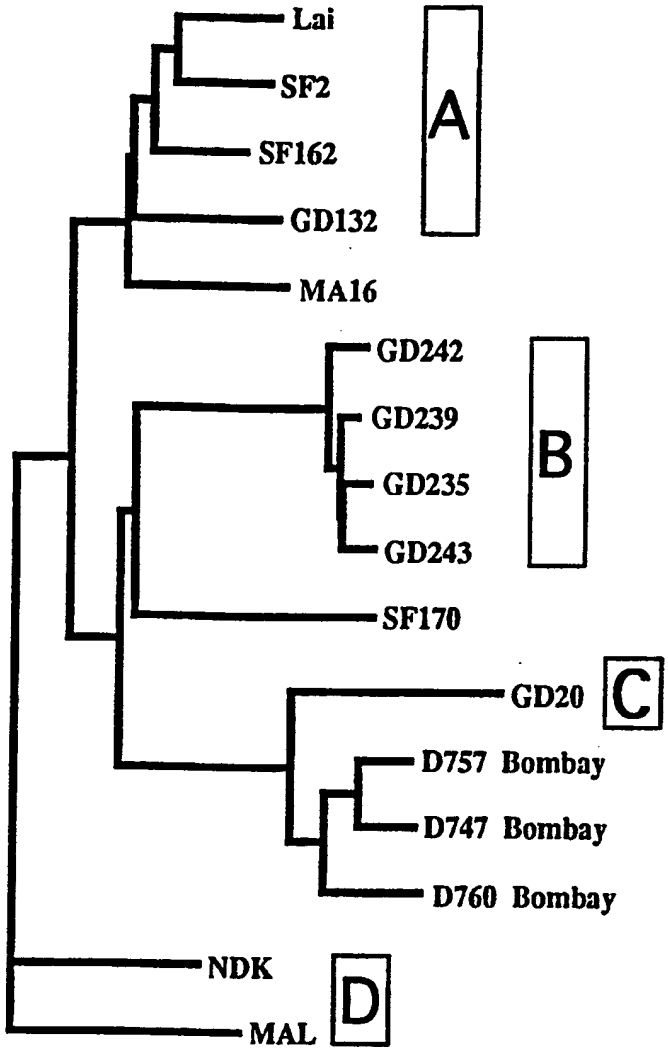

Lai
SF2
SF162
GD132
MA16
A
GD242
GD239
GD235
GD243
B
SF170
GD20
C
D757 Bombay
D747 Bombay
D760 Bombay
NDK
MAL
D

Fig. 9B

# INTERNATIONAL SEARCH REPORT

Interna al Application No
PCT/US 94/07416

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68 C12Q1/70

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | CURRENT OPINIONS IN BIOTECHNOLOGY, vol.3, January 1992, GB. pages 24 - 30 COTTON, R. ET AL 'detection of mutations in DNA' see the whole document --- | 1-25 |
| X | HUMAN GENETICS, vol.87, 1991, BERLIN DE. pages 728 - 730 CAI, S-P. ET AL 'a rapid and simple electrophoretic method for the detection of mutations involving small insertion or deletion: application to beta thalassemia' see the whole document --- -/-- | 1-25 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 22 November 1994 | 02 -12- 1994 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016 | Osborne, H |

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | MUTATION RESEARCH, vol.285, 1993, AMSTERDAM, NL. pages 125 - 144 COTTON, R. 'current methods of mutation detection' see page 132, paragraph 4 - page 133, paragraph 3 see page 137, paragraph 3 - page 139 --- | 1-25 |
| X | EP,A,0 443 748 (NATIONAL UNIVERSITY OF SINGAPORE) 28 August 1991 see the whole document --- | 1-25 |
| X | WO,A,90 13668 (LIFECODES CORP.) 15 November 1990 see page 7, line 30 - page 8, line 16 see page 9, line 10 - page 19 --- | 1-25 |
| Y | WO,A,91 00925 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY) 24 January 1991 see page 10, line 13 - page 14, line 7 --- | 1-25 |
| Y | WO,A,92 14844 (CALIFORNIA INSTITUTE OF BIOLOGICAL RESEARCH) 3 September 1992 see the whole document --- | 1-25 |
| Y | WO,A,93 08297 (BAYLOR COLLEGE OF MEDICINE) 29 April 1993 see the whole document --- | 1-25 |
| A | EP,A,0 405 376 (BOEHRINGER INGELHEIM INT. GMBH.) 2 January 1991 ----- | |

| Patent document cited in search report | Publication date | Patent family member(s) | | Publication date |
|---|---|---|---|---|
| EP-A-0443748 | 28-08-91 | AU-B- | 646655 | 03-03-94 |
| | | AU-A- | 7027691 | 08-08-91 |
| WO-A-9013668 | 15-11-90 | AU-A- | 5645690 | 29-11-90 |
| WO-A-9100925 | 24-01-91 | US-A- | 5045450 | 03-09-91 |
| | | CA-A- | 2062974 | 14-01-91 |
| | | DE-D- | 69008179 | 19-05-94 |
| | | DE-T- | 69008179 | 17-11-94 |
| | | EP-A- | 0482078 | 29-04-92 |
| | | JP-T- | 4506456 | 12-11-92 |
| WO-A-9214844 | 03-09-92 | AU-A- | 1662292 | 15-09-92 |
| WO-A-9308297 | 29-04-93 | AU-A- | 2931692 | 21-05-93 |
| | | CA-A- | 2121696 | 29-04-93 |
| | | EP-A- | 0610396 | 17-08-94 |
| EP-A-0405376 | 02-01-91 | CA-A- | 2019663 | 24-12-90 |
| | | DE-A- | 4020028 | 03-01-91 |
| | | JP-A- | 3117499 | 20-05-91 |
| | | US-A- | 5340713 | 23-08-94 |